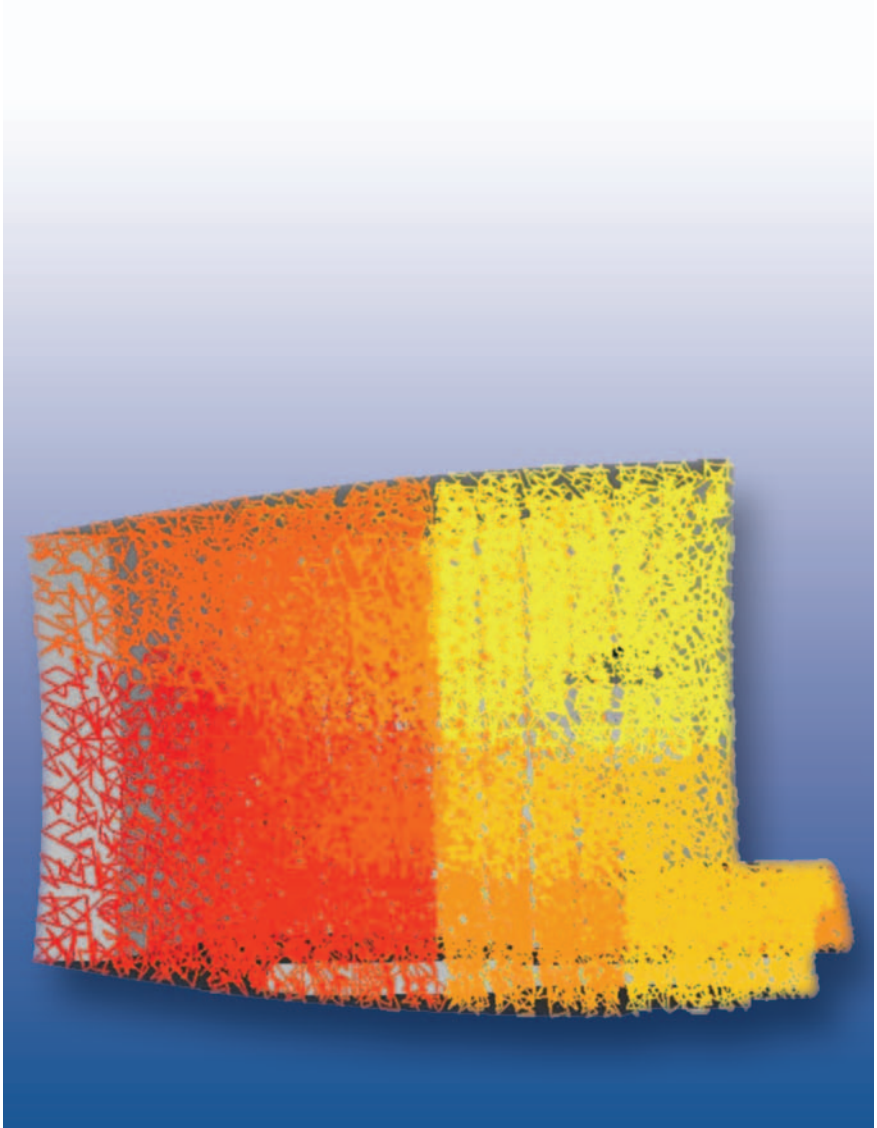
Summer Research Program

The ISCR put on its largest summer visitor program ever in FY 2004 with approximately 100 students in residence, as well as 11 faculty. Students are assigned individual LLNL mentors and given specific projects, ranging from programming tasks to original research, to which they will contribute based on their background and skills. Reports for most of the projects in the following table have been omitted from the printed version of the ISCR Annual Report in the interest of space, but can be found on the accompanying PDF. You can also obtain a CD-ROM containing these reports by calling (925) 423-3691.

To view the full abstract of any of the seminars listed on the following pages, simply click on the name of the seminar.



Summer Research Program

Name, Affiliation	Mentor, Host Organization
Project Title	

Sabbatical Visitors

Carl Ollivier-Gooch, University of British Columbia	Lori Diachin, CASC
Unstructured Mesh Technology Development	
Anne Greenbaum, University of Washington	Peter Brown, CASC
Boundary Integral Equation Methods and Software for Poisson's Equation in Three Dimensions	
Bjorn Sjogreen, Royal Institute of Technology, Stockholm, Sweden	Anders Petersson, CASC
Embedded Boundary Methods for Partial Differential Equations	

Summer Faculty

Constantin Bacuta, University of Delaware	Rob Falgout, CASC
A New Approach for Solving the Stokes Problem Based on the Distributed Relaxation Method	
James Brannick, University of Colorado at Boulder **Ludmil Zikatanov,** Pennsylvania State University	Rob Falgout, CASC Panayot Vassilevski, CASC
Compatible Relaxation and Optimal AMG Interpolation	
Zhiqiang Cai, Purdue University	Panayot Vassilevski, CASC
Mixed and Multigrid Methods for Stokes and Navier–Stokes Equations	
Tim Chartier, Davidson College	Van Emden Henson, CASC
Adaptive Multigrid via Subcycling on Complementary Grids	
Irene Livshits, University of Central Arkansas	Rob Falgout, CASC
AMG Algorithm for Finding an Eigenbasis for the Schroedinger Operator	
Anne Ngu, Southwest Texas University	Terence Critchlow, CASC
Large-Scale Integration of Web Sources	

Summer Students

David Alber, University of Illinois	Jim Jones, CASC
Adapting Algebraic Multigrid for the Solution of the Curl–Curl Formulation of Maxwell's Equations	
Kristopher Andersen, Unversity of California, Davis	John Pask, PAT
Locally Optimal Methods to Solve Eigenvalue Problems in Electronic Structure Calculations	
John Anderson, University of California, Davis	Benjy Grover, DCOM
QTester	
Dustin Anderson, California Polytechnic State University, San Luis Obispo	Don MacQueen, EPD
Real-Time Radiation Area Monitoring: Emergency Response and Regulation	
Benjamin Apodaca, Northern Arizona University	Carolyn Wimple, NIFE
PIMS Regression Test Development and Product Enhancement	

Name, Affiliation	Mentor, Host Organization
Project Title	
Abraham Bagherjeiran, University of Houston	Erick Cántu–Paz, CASC
Design and Implementation of an Anomaly Detector	
Paul Baginski, University of California, Berkeley	Van Emden Henson, CASC
Scalable Graph Algorithms	
Jeffrey W. Banks, Rensselaer Polytechnic Institute	William Henshaw, CASC
High-Speed Multicomponent Flows	
Lerone Banks, University of California, Davis	Jerry Rayome, CSP
Firewall Egress Filtering	
Janine Bennett, University of California, Davis	Valerio Pascucci, CASC
Volumetric Mesh Parameterization towards Slow-Growing Subdivision	
James Brannick, University of Colorado at Boulder	Rob Falgout, CASC
Adaptive Algebraic Multigrid Preconditioners in Quantum Chromodynamics	
Steve Callahan, University of Utah	Valerio Pascucci, CASC
A Memory-Insensitive Format for Out-of-Core Access to Unstructured Volumetric Meshes	
Jedidiah Chow, Granada High School	Jean Shuler, CSG
Discovery Center Display	
Kevin Chu, Massachusetts Institute of Technology	David Trebotich, CASC
Incorporating Electrokinetic Effects into the EB Navier–Stokes Embedded Boundary Incompressible Fluid Solver	
Kree Cole-McLaughlin, University of California, Los Angeles	Valerio Pascucci, CASC
Applying Morse Theory to Computational Datasets	**Terence Critchlow, CASC**
Dylan Copeland, Texas A & M University	Barry Lee, CASC
Geometric Multigrid for Variable Coefficient Maxwell's Equations	
Steven Davis, California State University, Hayward	Charles Doutriaux, EEBI
Web Designer, Program for Climate Model Diagnosis and Intercomparison	
Stanko Dimitrov, University of Arizona	Edmond Chow, CASC
Edge Betweenness Properties in Complex Networks	
Veselin Dobrev, Texas A & M University	Panayot Vassilevski, CASC
Element Agglomeration AMGe Solvers for Unstructured Finite-Element Problems	
Nina Dokeva, University of Southern California	Panayot Vassilevski, CASC
Parallel Implementation of a FETI–DP Method for Elliptic Problems with Mortar Finite Element Discretization	
Emily Eder, University of California, Los Angeles	Robert Fernandes, IOAC
Graph Viewer Improvements	

Summer Research Program

Name, Affiliation	Mentor, Host Organization
Project Title	
Christopher Egner, Rochester Institute of Technology	Paul Amala, AX Division
Platform Independence for Cale: wxWidgets and GNU Autotools	
Chris Elofson, University of Arizona	Vijay Sonnad, DCOM
Wavelet-Based Opacity Data Compression	
Dayo Esho, University of California, Berkeley	Steven Lee, CASC
ODE Visualization and Archive Tool in Python	
Pak-Wing Fok, Massachusetts Institute of Technology	Petri Fast, CASC
2D Numerical Modelling of Soap Film in Overture	
Camille Fournier, University of Wisconsin-Madison	Michael Kumbera, DCOM
Implementing Cray Pointers in the GNU gFortran Compiler	
Jeffrey L. Freschl, University of Wisconsin-Madison	John Johnson, DCOM
Performance Analysis of Monitoring and Information Systems Using NetLogger and an Investigation into P2P Overlay Topologies	
Evan Geller, Summit Preparatory High School	Albert Chu, HPSD
Bringing Order to CHAOS	
Tobias Gradl, Technische Universität München, Germany	Edmond Chow, CASC
Assessing Performance of Hybrid MPI/OpenMP Programs on SMP Clusters	
Rachel Greenstadt, Harvard University	David Youd, NAIC
Censorship Resistance	
Attila Gyulassy, University of California, Davis	Valerio Pascucci, CASC
Hierarchical Morse–Smale Complexes	
Daniel Han, University of Southern California	Terry Jones, ICC Services and Development Division
MPI Profiling	
Kevin Hoffman, Brigham Young University	Robert Cooper, DCOM
Parallel Analysis of Asymmetries in Symmetric Simulations in ALE3D	
W. Taylor Holliday, University of California, Davis	Valerio Pascucci, CASC
Combinatorial Feature Extraction for a Streaming Framework	
Gary Hon, University of California, San Diego	David Clague, EE-EETD
Virtual PCR (vPCR)	
Jason Howell, Clemson University	Carol Woodward, CASC
A Two-Grid Method for Radiation Diffusion	
David Hoyt, Brigham Young University	James Schek, NAIC
DHS Countermeasures Test Bed–Command Center	

Name, Affiliation	Mentor, Host Organization
Project Title	
Kevin Huck, University of Oregon	Brian Miller, CASC
PerfTrack	
Lukas Jager, Universität Bonn, Germany	Radu Serban, CASC
A Multilevel Preconditioner for a PDE-Constraint Optimization Problem	
Abram Jujunashvili, University of Colorado at Denver	Charles Tong, CASC
Integration of Multivectors to Hypre	
Alin Jula, Texas A&M University	Dan Quinlan, CASC
Automatic Detection of Hot Spots in C/C++ Programs	
Kirk Kelsey, University of Rochester	Tom Epperly, CASC
Automatic SIDL Generation with ROSE	
Jason Kimball, University of California, Irvine	Mark Duchaineau, CASC
Large-Scale Atom Data Visualization	
Nicholas Kridler, University of Colorado at Boulder	Barry Lee, CASC
Algebraic Multigrid for Maxwell's Equations with Variable Coefficients	
Brian Lum, University of California, San Diego	Amber Marsh, EEBI
IMAGE Pipeline	
Eric Machorro, University of Washington	Britton Chang, CASC
Numerical Solution Methods and Error Analysis for the Neutron Transport Equation in Absorbing, Monoenergetic, Non-Scattering Media	
Anna Majkowska, University of California, Riverside	Valerio Pascucci, CASC
Out-of-Core Visualization of Climate Modeling Data	
Robert Dean Malmgren, Northwestern University	Scott Brandon, AX Division
Code-Independent Analysis Tools for Physics Simulation Codes	
Peter Manning, University of Arizona	Glenn Goderre, NIFE
Development of a Graphical User Interface for the Virtual Beam Line (VBL) Simulation Software	
Ajith Mascarenhas, University of North Carolina at Chapel Hill	Valerio Pascucci, CASC
Time-Varying Reeb Graphs for Space-Time Data	
Tammara Massey, University of California, Los Angeles	Terry Brugger, NAIC
Ultra-Wideband Token Ring Simulation and Security in Sensor Networks at LLNL	
Kathryn Mohror, Portland State University	John May, CASC
PerfTrack	
Christopher Muelder, University of California, Davis	Marv Christensen, NAIC
Summarizing Network Traffic with Information Visualization	

Summer Research Program

Name, Affiliation	Mentor, Host Organization
Project Title	
Spencer Nielsen, Brigham Young University	Darrel Whitney, CADSE
Software Failure Risk Analysis	
Daniel Oeltz, Universität Bonn, Germany	Panayot Vassilevski, CASC
A Space-Time Sparse Grid Approximation Space	
Christopher Olson, University of California, Santa Cruz	Robert Cooper, DCOM
Reviewing Vista	
Timothy Paik, University of California, Berkeley	Shawn Larsen, ISCR Dave Harris, ISCR
Subspace Detectors	
Sung W. Park, University of California, Davis	Peter Lindstrom, CASC
Streaming Pointsets	
Bryan Parno, Harvard University	Tony Bartoletti, NAIC
Internet Ballistics: Retrieving Forensic Data From Network Scans	
Dan Phung, Columbia University	Moe Jette, CASC
SLURM for BlueGene/L	
Sriram Polepeddi, Carnegie Mellon University	Noel Tijerino, IOAC
Software Vulnerability Taxonomy Consolidation	
Raymond Pon, University of California, Los Angeles	Terence Critchlow, CASC
Performance-Oriented, Privacy-Preserving Data Integration	
Serban Porumbescu, University of California, Davis	Mark Duchaineau, CASC
Out-of-Core Parameterization and Remeshing of Surfaces	
Ashley President, Carnegie Mellon University	Michael Kumbera, A Division
Intrinsic Function Testing and an OpenMP Runtime Library Summer	
Davinder Rama, California State University, Sacramento	Adam Zemla, EEBI
Amino-Acid Sequence into Tertiary Structure Web Site	
George Roberts, Georgia Institute of Technology	Chandrika Kamath, CASC
Variance-Based Feature Tracking	
Rolf Ryham, Penn State University	Rob Falgout, CASC
Multigrid Prolongation Based on Sharp Convergence Theory	
Andreas Saebjornsen, Universitete i Oslo, Norway	Dan Quinlan, CASC
The AST Query Mechanism and the C/C++ Graphing Mechanism	
Elmer Salazar, California State University, Stanislaus	Michael Kumbera, DCOM
Implementation of OpenMP Support in the gFortran Compiler	
Jennifer Sirp, California State University, Sacramento	Terry Brugger, NAIC
Extending ReiserFS for Automatic File Queuing	

Name, Affiliation	Mentor, Host Organization
Project Title	
Yoshifumi Suzuki, University of Michigan	Jeff Hittinger, CASC
Efficient Schemes for Hyperbolic Systems with Stiff Relaxation Source Terms	
Valerie Szudziejka, University of California, Davis	Valerio Pascucci, CASC
Streaming Computation of Structural Graphs	
Ryan Szypowski, University of California, San Diego	Ulrike Yang, CASC
Parallel AMG for Systems of PDEs	
Brian Taylor, University of Illinois at Urbana-Champaign	Bill Henshaw, CASC
Computation of Cellular Detonation	
Robert Taylor, Northern Arizona University	Jody Malik, CSP
Vulnerability Tracking Database 2.0	
Nils Thuerey, Universität Erlangen-Nürnberg, Germany	Dan Quinlan, CASC
Global Analysis of the ROSE Infrastructure	Jody Malik, CSP
Peter Tipton, University of Southern California	Brandon Scott, AX Division
Code Validation Made Easy	
Chunbo Wang, Purdue University	Charles Tong, CASC
A MATLAB Implementation of Mixed Finite-Element Method for Incompressible Newtonian Flows: Pseudostress–Velocity Formulation	
Rebecca Wasyk, Worcester Polytechnic Institute	Carol Woodward, CASC
Newton–Krylov Methods for Expensive Nonlinear Function Evaluations	
Sage Weil, University of California, Santa Cruz	Tyce McLarty, CADSE
Metadata Management for Petabyte-Scale File Systems	
Daniel Wendlandt, Stanford University	Martin Casado, NAIC
Information Leakage Due to Geographic Properties of Internet Routing	
Brian White, Cornell University	Dan Quinlan, CASC
Ameliorating the Performance Degradation of User-defined Abstractions and Indirect Memory Accesses	
Ryan M. White, University of California, Berkeley	Shawn Newsam, CASC
Matching Shapes Using Local Descriptors	
Jeremiah Willcock, Indiana University	Dan Quinlan, CASC
Additions to the ROSE Compiler Infrastructure	
Suzanne Wingenter, San Diego State University	Petri Fast, CASC
Shallow-Water Equations on Curvilinear Grids	
Christopher Wojtan, University of Illinois, Urbana-Champaign (2000–2004) Georgia Institute of Technology (2004)	Jeremy Meredith, DNT
A Hybrid Sort-First/Sort-Last Approach for Rendering Translucent Geometry in the VisIt Visualization Tool	

Summer Research Program

Name, Affiliation	Mentor, Host Organization
Project Title	
Sung-Eui Yoon, University of North Carolina at Chapel Hill	Peter Lindstrom, CASC
Cache Coherent Mesh Layout	
Beth Yost, Virginia Polytechnic Institute and State University	Terence Critchlow, CASC
A Visual Interface for the Promoter Identification Workflow	

Unstructured Mesh Technology Development

Student

Carl F. Ollivier-Gooch, University of British Columbia

Mentor

Lori Diachin, CASC

The research proposed under this subcontract consisted of three parts:

1. Enable Mesquite to use GRUMIVW topology modification routines and enable GRUMMP to use Mesquite's new vertex smoothing technology
2. Explore possibilities for globally optimal mesh reconnection.
3. Attempt to create a high-order accurate unstructured mesh version of a computational electromagnetics code, and explore the possibilities for adding high-order accurate unstructured mesh capabilities to Overture.

This report describes the work done on each of these three topics and future prospects in each area.

1. Mesquite-GRUMMP Interconnection

This task decomposes naturally into two sub-tasks: enabling communication of mesh data between two large pieces of unstructured mesh manipulation code that use very different data structures, and supplying the small amount of control information to specify what the mesh improvement goal is.

Communication of mesh data between GRUMMP and Mesquite is simplified, or at least given a specific structure, by the Terascale Simulation Tools and Technologies (TSTT) mesh interface definition. While at LLNL, I implemented the TSTT mesh interface using the GRUMMP mesh database as the back end. In the process, I also identified several missing or ambiguous items in the interface and made proposals to the TSTT interface group to correct them. Major issues that I have helped to clarify in the TSTT interface include error handling and entity sets. At this point, the GRUMMP implementation of the TSTT interface, while not yet 100% complete, is more than sufficient to support all operations required by Mesquite for vertex smoothing.

The reverse communication – Mesquite implementation of the TSTT interface – is not yet complete.[1] By the time it is, implementation of mesh reconnection using the algorithms currently in GRUMMP but with mesh changes through the TSTT interface should also be complete. This will enable not only mesh reconnection in Mesquite but in other TSTT-based tools, as well. Work on the TSTT mesh reconnection service has already begun.

Once full two-way mesh data communication between Mesquite and GRUMMP is possible, implementing GRUMMP's mesh reconnection in Mesquite and Mesquite's new vertex smoothing in GRUMMP will require little more than specifying the quality measure to optimize and calling the optimization routine.

2. Global Mesh Reconnection

The goal of this part of the project was to determine the feasibility of using global rather than local mesh reconnection techniques to improve unstructured mesh topology. I began by creating a naive implementation of the well-known branch-and-bound strategy for combinatorial optimization. The initial implementation demonstrated that global reconnection is possible, but execution times were absurdly high. A series of improvements to the algorithm were made, all aimed at improving the effectiveness of branch pruning, because effective pruning can dramatically reduce the number of possible mesh configurations that must be examined in determining the best global connectivity. As a result of these efforts, the original code was sped up by a factor of over 1000 for small meshes. Unfortunately, run times are still relatively slow, and the trend in run time with mesh size is poor: in three dimensions, run time scales roughly with the seventh power of mesh size, which is clearly prohibitive. Nevertheless, it was felt that the ideas involved were of sufficient interest to the meshing community to justify a conference paper, and a draft paper was submitted to the 13th International Meshing Roundtable. (Attached)

[1] Mesquite coding will be done by a DOE employee or post-doc, so timing on that is out of my control.

Carl F. Ollivier-Gooch, (Continued)

3. High-order Discretization

The main effort here was in discussions with the Overture team about the possibility of adding high-order unstructured operators to the Overture framework. I learned how Overture is constructed and how it provides various differential operators efficiently at high orders of accuracy for structured meshes. I also discussed at some length with Bill Henshaw and Kyle Chand of the Overture team how one might implement these types of operators using the high-order unstructured mesh solution reconstruction techniques that my solver library provides. While we all agreed that using my high-order reconstruction code to extend Overture's capabilities is both useful and feasible, we reached no firm conclusions about how and when to proceed with implementation.

Boundary Integral Equation Methods and Software for Poisson's Equations in Three Dimensions

Student

Anne Greenbaum, Professor of Mathematics, University of Washington

Eric Machorro, Graduate Student, University of Washington

Mentor

Britton Chang, CASC

During my sabbatical quarter at LLNL, I worked on two projects. One involved the fast solution of Poisson's equation on arbitrary 3D regions, and the other involved finding new methods and analyzing current methods for solving the linearized Boltzmann equation for neutron transport.

For the first project, I implemented a code for solving Poisson's equation on an arbitrary 3D region by first embedding the region in a cube. Next, I solved a boundary-integral equation to determine the right-hand side vector for a fast Poisson solver and then applied the fast Poisson solver on the cube. This algorithm had been described in the literature (1) but had not previously been implemented as a package. In doing so, I unfortunately discovered a few drawbacks.

First, the boundary integral equation must be solved to greater than second-order accuracy in order to obtain second-order accuracy in the Poisson solver. This can be done using a form of Richardson extrapolation, but it means that the user must supply more than a simple triangulation of the surface. Next, while the boundary integral equation can be solved with just a few GMRES iterations, the cost of forming the matrix and multiplying the matrix at each iteration is great. To make the method competitive, these matrix-vector multiplications need to be carried out with the fast multiple method. Finally, an improved procedure for calculating boundary values on the cube needs to be implemented. The conclusion was that the current code is not yet competitive with other methods, such as multigrid methods, for solving Poisson's equation on interior regions, although it might be the method of choice for exterior problems where finite difference methods may not be applicable.

The other project I worked on, along with graduate student Eric Machorro, had to do with neutron transport. We were particularly interested in methods that would avoid negative fluxes and oscillations, yet would maintain second-order accuracy or better. Eric implemented the exponential characteristic (EC) method (2), which avoids negative fluxes and maintains conservation but is considerably more expensive per cell than the more standard diamond-difference or Petrov-Galerkin methods. Moreover, we were able to show that it is only first-order accurate in the mathematical sense, although it may obtain high accuracy for certain problems anyway. Working with Britton Chang of LLNL, we were able to rigorously establish second-order accuracy for the Petrov-Galerkin method, a property that had previously been suggested but never proved. A paper on this has been written and will soon be submitted for publication.

References

1. Greenbaum and Mayo, "Rapid Parallel Evaluation of Integrals in Potential Theory on General Three-Dimensional Regions," *Jour. Comp. Phys.*, 145 (1998), pp. 731-742

2. Minor and Mathews, ``Exponential Characteristic Spatial Quadrature for Discrete Ordinates Radiation Transport with Rectangular Cells," *Nuclear Sci. Engr.*, 120 (1995), pp. 165-186

Embedded Boundary Methods for Partial Differential Equations

Student

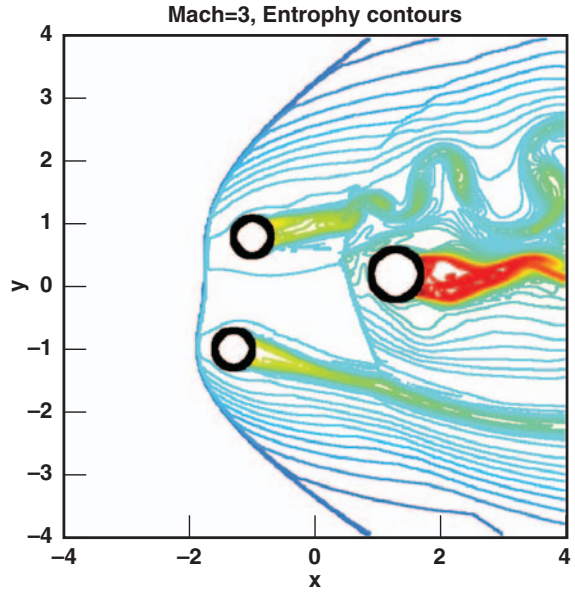
Bjorn Sjögreen, Royal Institute of Technology, Stockholm, Sweden

Mentor

Anders Petersson, CAR, CASC

First, we investigated far-field boundary conditions for Maxwell's equations. After a preliminary study of different types of boundary conditions, we decided to use the perfectly matching layer (PML) technique. Although this is a well-known method, there are several parameters in the equations that are usually selected based on computational experience. In my work, I have been able to show mathematically exactly how these parameters should be chosen in order to have a grid convergent method. This far-field boundary condition was implemented to compute scattering from perfectly conducting objects. The objects were represented as embedded curves in a Cartesian grid using the new embedded boundary method by H.O. Kreiss and A. Petersson. As a second project, I redefined the above embedded boundary method for Maxwell's equations to be able to solve compressible fluid flow problems.

The original embedded boundary method did not perform well for such problems because of the shockwaves that usually exist in compressible fluids. It was therefore necessary to modify the embedded boundary interpolation formulas. I introduced special limiter functions into the boundary interpolation and narrowed the interpolation stencil. With these changes, the method performs very well. An example of flow past three disks at Mach 3 is shown below. Finally, I have implemented the method for three space dimensions and computed compressible fluid flow past a sphere.



The figure shows streamlines for air flowing in from the left at supersonic speed. A strong bow shock is formed in front of the cylinders, a smaller shock can be seen in front of the rearmost cylinder.

TOC

A New Approach for Solving the Stokes Problem Based on the Distributed Relaxation Method

Student
Constantin Bacuta, University of Delaware

Mentor
Rob Falgout, CASC
Panayot S. Vassilevski, CASC

We present a distributed relaxation method for the incompressible Stokes problem. The method can be applied to discretizations of the Stokes system provided that the discrete approximation spaces for velocity and for pressure form a stable pair. We establish sharp rates of convergence in terms of the inf-sup condition constants at the continuous and the discrete level. We prove that our iterative process is optimal among all Uzawa-type algorithms that involve constant step-size parameters. Based on the functional iteration formulation of the new algorithm, we are able to analyze Achi Brandt's distributed relaxation method as an iterative process and compare it with Uzawa-type methods for solving the Stokes system. The distributed relaxation method for the incompressible Stokes problem is based on a change of variables that leads to a lower triangular system with Laplace operators on the main diagonal for which multigrid methods are more suitable.

We propose a finite-element formulation based on the distributed relaxation method. Using the new approach, we can construct effective and robust multigrid methods for solving Stokes-type systems and other PDE elliptic systems that can be reformulated as saddle-point problems.

We plan to present numerical experiments to demonstrate the effectiveness of the transformation, which is well-established for certain finite difference discretizations of Stokes problems.

For a future research activity, we consider the Maxwell equations, reformulated as saddle point problem. With an appropriate change of variables, a similar iterative process can be used to obtain an efficient way to discretize and solve the Maxwell equations.

References

Bramble, J.H., J.E. Pasciak and P.S. Vassilevski, "Computational scales of Sobolev norms with application to preconditioning," *Mathematics of Computation*, 69 (2000), pp. 463-480.

Brandt, A., and N. Dinar, "Multigrid solutions to flow problems," *Numerical Methods for Partial Differential Equations,(S. Parter, ed.),* 1979, Academic Press, New York, pp. 53-147.

Girault, V. and P.A.Raviart, "Finite Element Methods for Navier-Stokes Equations,"
Springer-Verlag, Berlin, 1986.

Wittum, G., "On the convergence of multi-grid methods with transformed smoothers,"
Numerische Mathematik, 57 (1990), pp. 15-38.

TOC

Compatible Relaxation and Optimal AMG Interpolation

Student

James Brannick, University of Colorado, Boulder

Ludmil Zikatanov, Pennsylvania State University

Mentor

Rob Falgout, CASC

Panayot Vassivlevski, CASC

In collaboration with Rob Falgout and Panayot Vassivlevski, we worked on the development and implementation of some state-of-the-art techniques for the construction of Algebraic Multigrid Methods (AMG) for the solution of positive definite linear systems arising from the discretizations of elliptic partial differential equations. Our approach uses the fact that a Compatible Relaxation (CR) process reduces the condition number of the matrix corresponding to the fine grid (usually denoted with Aff) as much as possible. Our idea is that the supports of the coarse grid basis functions (Ωi) can be determined by constructing an approximation to Aff-1Afc (Afc represents the coupling between coarse-grid and fine-grid degrees of freedom). Once the supports are defined, we determine the basis in a way that the trace of the coarse grid matrix is as small as possible.

Such a scheme can be made quite robust, by dynamically taking better and better approximation to Aff-1Afc. This is done by increasing the size of Ωi or updating the approximation to the slow-to-converge error, which is obtained by applying the iteration to the homogeneous system of equations Au = 0 and then and recomputing the basis during the V-cycle iterations. Our current studies show that a well-conditioned Aff, provided by a CR procedure, leads to a good approximation, even by choosing Ωi to be a small size.

The set of the test problems we have used so far is on matrices corresponding to discretizations of scalar PDEs. Our plan is to extend these techniques and implement them in the case of vector equations and systems of PDEs, such as linear elasticity and Maxwell equations. Although it may not be optimal, we are confident that an ad hoc application of the idea described above will work in these more complicated cases.

TOC

Mixed and Multigrid Methods for Stokes and Navier-Stokes Equations

Student

Zhiqiang Cai, Purdue University

Mentor

Panayot Vassilevski, CASC

To discretize the incompressible Stokes and Navier-Stokes equations by the finite element method, one requires a stable combination of finite element spaces in order to approximate the velocity and pressure in a well-defined and optimal manner. Commonly used stable pairs include the Taylor-Hood element. The algebraic equations resulting from the discretization is symmetric but indefinite. Many solution methods that work well for symmetric and positive definite problems cannot be applied directly. Although substantial progress in solution methods for saddle-point problems has been achieved, these problems may still be difficult and expensive to solve.

The purpose of this project is to develop, analyze and implement accurate and efficient numerical methods for solving the Stokes and Navier-Stokes equations. Instead of using the velocity-pressure formulation, our method is based on the pseudostress-velocity formulation developed jointly with LLNL scientists. The pseudostress is approximated by the Raviart-Thomas (RT) and the velocity by piecewise discontinuous polynomials. We have shown that this pair of finite elements is stable and yields quasi-optimal accuracy. Important physical quantities, such as the pressure, vorticity and stress, may be calculated algebraically through post-processing.

To efficiently solve the indefinite system of linear equations resulting from the discretization of the stationary Stokes equation, we decouple the system by the penalty method. The penalty parameter is chosen to be proportional to the approximation accuracy of the discretization for avoiding accuracy loss. Even though the reduced pseudostress system is very ill-conditioned, we have developed an efficient multigrid method based on the H(div)-type of multigrid method. The velocity can then be recovered without solving any equations. Numerical experiments on uniform rectangular RT elements of the lowest order show that multigrid convergence factor is around 0.25 and that it is independent of the mesh size, the number of levels, and the penalty parameter.

In the future, we will study efficient multigrid methods for solving system of linear equations resulting from the discretization of the time-dependent Stokes equations. This is very challenging because the resulting pseudostress system at each time step has many extra near-null space components. We will also continue our study for both stationary and time-dependent Navier-Stokes equations.

TOC

Adaptive Multigrid via Subcycling on Complementary Grids

Student **Timothy Chartier,** Davidson College

Mentor **Edmond Chow,** CASC

Algebraic multigrid methods uncover or assume the nature of algebraically smooth error components in order to construct multigrid components. Adaptive multigrid schemes expose algebraically smooth error, analyze the effectiveness of the resulting multigrid algorithm and adjust the cycling as needed in order to improve the rate toward convergence.

This research, conducted in collaboration with Edmond Chow, focused on the use of relaxation and subcycling on complementary grids as an evaluative tool in correcting multigrid cycling. Each complementary grid is constructed with the intent of dampening a component of algebraically smooth error. In particular, complementary grids are constructed when the adaptive cycle determines that smooth modes are not being sufficiently eliminated by the current multigrid process. The particular implementation of this idea manages smooth error in a manner analogous to spectral AMGe.

The algorithm was tested on a model problem with Dirichlet boundary conditions. On both the isotropic and anisotropic Laplacian operators, convergence rates for 2-4 levels remained at less than 0.10. It is important to note that comparable results occurred when the system $Ax = b$, was scaled to produce the new system $SASx = b$

where S was a diagonal matrix with $s_{ii} = 10^{d_i}$ and $0 < d_i < 103$. For the rotated anisotropic diffusion problem, convergence rates were kept below 0.20 for 2-4 levels. Again, the scaling, which troubles traditional algebraic multigrid, produced comparable results.

Future numerical tests will include larger systems and a more extensive pool of problems.

TOC

AMG Algorithm for Finding an Eigenbasis for the Schroedinger Operator

Student

Irene Livshits, University of Central Arkansas

Mentor

Robert Falgout, CASC

The goal of this project is to develop multigrid solvers that allow an efficient approximation to eigenvalues and eigenfunctions of Schroedinger operators with variable potentials. The developed method allows fast and inexpensive calculation of all eigenvalues in a given interval, as well as the corresponding eigenfunctions. The calculated eigenbasis has a multilevel structure – the number of eigenfunctions approximated on each grid grows as grids become coarser. All eigenfunctions are approximated and the eigenvalues are calculated only on the coarsest grid. This structure is not only inexpensive to calculate and store, it also allows for efficient use of the eigenbasis, e.g. for a fast summation.

The algorithm deals with the indefinite nature of a Schroedinger operator using the wave-ray approach developed by Livshits and Brandt for Helmholtz-boundary value problems with constant potentials. However, unlike the geometric wave-ray method, the new algorithm employs algebraic multi-Galerkin approach to construct coarse grid operators. To date, the solvers were developed for one-dimensional operators, serving as a calibration for two- and three-dimensional algorithms.

The suggested approach not only aims to solve computationally challenging Schroedinger equations, it can be also used for other operators that have several independent low eigenfunctions.

TOC

Large-scale Integration of Web Sources

Student
Anne Ngu, Texas State University

Mentor
Terence Critchlow, CASC

I worked on two specific aspects of this project during this summer. The first aspect is a continuation of last summer's work in the automatic discovery and classification of Web sources. My task is to apply the metadata-based classification methodology used for identification of BLAST servers to other application domains. This involves creating a service class description for bio-keyword search applications and conduct experiments to measure the performance of our classifier for this domain. The result is incorporated in the paper submitted to the World Wide Web Journal.

The second aspect of my work this summer was to investigate the possibility of automatically generating a service class description for automatic classification. The manual creation of service class description is tedious, error prone and not adaptative to the dynamic nature of Web sources. The idea is to use a supervised learning technique for automatically generating the regular expressions that can be used for recognizing the data embedded in an HTML response page. Given training documents and a set of user-annotated examples, including a few sample sites of a class of Web source, the system will automatically generate a set of data type rules in the form of regular expressions that can be used for the correct identification of all the sources that belong to that class.

I implemented a clustering-based learner algorithm for automatic generation of data type rules and incorporated it with the existing dataRecognizer system at LLNL. The initial experimental evaluation of the generated rules showed them to be superior with respect to the quality of rules and document classification time. A report entitled "Automatic Generation of Data Types for Classification of Deep Web Sources" is being prepared.

Adapting Algebraic Multigrid for the Solution of the Curl-Curl Formulation of Maxwell's Equations

Student **David Alber,** University of Illinois at Urbana-Champaign

Mentor **Barry Lee,** CASC

Applying multigrid to solve linear systems arising from the discretization of Maxwell's equations has been a difficult problem largely because of the large null space of the curl-curl operator. Other solvers, such as Krylov subspace solvers, will work on such systems. However, the desire to have a solver that will scale effectively on a larger numbers of processors is a great motivation to adapt multigrid to the task. With more than 65,000 processors, BlueGene/L is a good architecture to highlight the need for scalable solvers.

Multigrid methods have been created to solve Maxwell's problems, but these methods tend to degrade in performance when the values of the coefficients in the problem vary. Jim Jones and Barry Lee recently introduced a method designed to handle these more complicated problems effectively by allowing the curl- and divergence-free near-null space error components to be annihilated separately. The method shown in their paper is for use on two-dimensional problems on structured grids. Its viability on three-dimensional structured problems was shown by Dylan Copeland.

The purpose of this work was to begin creating a solver based on the same principles that work in two-dimensions on problems defined on unstructured grids. This summer, tests were run in an attempt to approach this problem in different ways using an algebraic multigrid. It is clear that more sophisticated ways of addressing the problem using algebraic multigrid need to be investigated. In particular, an effective way to select coarse grids for curl- and divergence-free errors needs to be formulated.

Locally Optimal Methods to Solve Eigenvalue Problems in Electronic Structure Calculations

Student **Kris Andersen,** University of California, Davis

Mentor **John Pask,** H Division: Metals and Alloys Group

The goal of my project was to find the most efficient way to implement a new locally optimal method to solve generalized eigenvalue problems – developed by Andrew Knyazev and Richard Lehoucq within the context of electronic structure calculations. Since the matrices involved are large (106) and the tolerances are more stringent than typical engineering applications, this involved paying close attention to compromises between efficiency and robustness. Specifically, I developed a new way to handle deflation that makes use of the reasonable eigenvector approximations available in an electronic structure calculation. I performed a number of numerical experiments to find the most efficient way to do specific operations, such as restarting the conjugate gradient algorithm. I also looked closely at preconditioning and learned how to interface our code (which uses complex arithmetic) with the HYPRE package (which only supports real numbers). I also explored how to use numerically cheaper preconditioners, such as a PCG linear solver, more effectively.

TOC

QTester

Student **John Anderson,** University of California, Davis

Mentor **Benjy Grover,** DCOM

This summer, I rearchitected the QTester application, a graphical user interface testing tool. QTester makes it easy to design and run regression tests via the graphical user interface of any Qt-based application. Users can create test cases while using their target application. The QTester works by recording Qt events, such as mouse clicks and key presses, and then replaying the events upon request. For most applications, the QTester opens up greater opportunities for assuring software quality.

Real-Time Radiation Area Monitoring: Emergency Response and Regulation

Student **Dustin Anderson,** California Polytechnic State University, San Luis Obispo

Mentor **Don MacQueen,** Environmental Protection Department

Sixteen Geiger-Mueller sensors set up around the Lawrence Livermore National Laboratory continuously monitor for harmful radiological conditions. Using meteorological data from the Lab's meteorological tower, the Real-Time Radiation Area Monitoring Network (RTRAM) compares current radiation counts to reference values to determine threat severity. The networked sensors on site are connected to a Central Command Center where their radiation readings are rerouted in real-time. Incorporated into this network is an electronic notification system that immediately alarms the Emergency Operations Center about hazardous conditions. If a situation arises, the threat is automatically reported to the EOC so that appropriate measures may be taken in a timely manner. The RTRAM network's purpose at the Lab is twofold; it doubles as both a safety net for the Lab and ensures that daily radiation emissions are within regulations.

The software managing RTRAM's web site is written in Perl, SQL, and now JavaScript. My participation in the project included working on these scripts to optimize how data is retrieved from the database, how the data is organized and formatted in the real-time viewable web pages, and how the data is stored, adding functionality, and fixing compatibility issues while converting old data to new data. This new means of storing data improves real-time performance, which increases usability and improves the efficiency of the entire RTRAM network. Boosting performance allows users to see data faster than ever and helps them respond to critical situations more quickly.

Related Projects

In order to be better prepared to renovate the RTRAM network and make my work on it as efficient as possible, I took on several side projects to help me better understand the programming languages Perl, JavaScript and SQL, as well as the nuances of HTML and XML. The other projects I worked on as pseudo-training before tackling the RTRAM network software included, but are not limited to:

Oracle Password Pages

The first project that introduced me to the Perl programming language and also to database communication using Perl and SQL was a set of two pages that enabled users to change their password on the Oracle database. Users would enter all relevant information and after error-checking all entered data, the user's password would be changed as long as all tests were passed. This was also my first exposure to SQL, a language designed for database communication through the use of queries.

QC Chemist Validation Web Pages

These Quality Control (QC) Web pages were written as a replacement for a tool on the old Ingres server used by the ERD ISMG chemists. Users can now access all chain of custodies (COCs) on the Oracle database by entering a COC ID and then altering and adding data as necessary. These pages, and the DMT Receive Data Web Pages described below, were critical milestones in ERD ISMG's target goal of going parallel and switching from the Ingres database system to the Oracle database system.

Data Management Tools (DMT) Receive Data Web Pages

Similar to the QC Chemist Validation pages, the DMT Receive Data pages allow data managers to access different aspects of all chain of custodies on the Oracle Database by entering a desired COC ID. Information regarding a specific COC can be altered and/or deleted through these pages, allowing data managers to both double-check and enter new queries.

CES Algorithm Review Web Pages

The Cost Effective Sampling (CES) Algorithm Review Web pages are used by the Environmental Restoration division of the Environmental Protection Department to help determine how often it samples ground water monitoring wells. Altering these pages included adding automatic email verification and redirection of pages, as well as formatting the pages

Dustin Anderson (Continued)

to be consistent with the other data management tools (DMTs).

Drop Tables Pages

Enhancing the previous "Drop Tables" Web page, the new Drop Tables pages allow data managers to drop tables from the Oracle database. Previously, users had no choices to make, but the new pages allow users to drop either individual tables or all tables at once. This was my first encounter with JavaScript, and I used these skills later when I altered the RTRAM network scripts.

The Terrestrial and Atmospheric Monitoring and Modeling Group Web Page

This was my main training in using HTML and XML to format data. Creating a web page with the use of explicit table layouts, the TAMM group's overview web page is now both informative and aesthetically pleasing.

Documentation of the RTRAM Web Site

In order for others who join the RTRAM project to have an easier time learning, understanding and navigating the network, I created 20 separate Web pages written in HTML/XML and linked together as an "Electronic Handbook." The Handbook can be browsed as a normal Web site, and each page summarizes exactly one Perl script that runs the site with links mimicking those on the actual site. This document provides summaries of each of the Perl scripts, along with showing how each script relates to all the others.

Query By Form (QBF) Tool: Group Delete Functionality

Adding functionality to the already existing QBF tool required me to add elements to the Web tool in order to make the data managers' jobs a little easier. Basing my code on the Drop Tables pages mentioned above, I enabled group delete functionality to this tool, giving data managers the ability to delete specific rows from their data tables as individuals or as user-defined groups.

Adding New Functionality to the RTRAM Website

Although many features were altered and/or optimized during "renovation," two significant new features and four Perl scripts were added. Comments can now be added to censor data during specific times and/or time intervals. These comments can also be easily viewed through the Web site – features that were not previously offered.

Converting Old Data & Compatibility

Old data, stored in unusable format, can now be converted to the new data format easily with a Perl script. Any number of data files in the old format can be converted to the new format simultaneously, enabling data retrieval via the Web to produce a *complete* data set, rather than the partial set that was previously created.

PIMS Regression Test Development and Product Enhancement

Student **Benjamin Apodaca,** Northern Arizona University

Mentor **Carolyn Wimple,** NIFE

This summer I worked on three main tasks: creating testing scripts to perform regression testing on a space utilization system called Engineering Facility Information (EFI), creating a prototype of the Skills section of the Personnel Information Management System (PIMS), and creating a Web version of the EFI User Instruction document.

I was introduced to a software tool called WinRunner, which I had never used before, and used it to create the test scripts for EFI. I was also introduced to HTML and JavaScript, and I used these Web programming languages to create both the Skills prototype and the Web version of the EFI User Instruction document. All three of the tasks are part of a larger project focused on the PIMS, which is still being performed by the team on which I worked.

Design and Implementation of an Anomaly Detector

Student **Abraham Bagherjeiran,** University of Houston

Mentor **Erick Cantú-Paz,** CASC

I designed and implemented a general-purpose anomaly detector for streaming data. Based on a survey of similar work from the literature, a basic anomaly detector builds a model on normal data, compares this model to incoming data, and uses a threshold to determine when the incoming data represents an anomaly. Models compactly represent the data but still allow for effective comparison.

Comparison methods determine the distance between two models of data or the distance between a model and a point. Threshold selection is a largely neglected problem in the literature, but the current implementation includes two methods to estimate thresholds from normal data.

With these components, a user can construct a variety of anomaly detection schemes. The implementation contains several methods from the literature. Three separate experiments tested the performance of the components on two well-known and one completely artificial dataset. The results indicate that the implementation works and can reproduce results from previous experiments. For more details, see the technical report on the Sapphire Web site at http://www.llnl.gov/casc/sapphire/

I also update and repaired problems with the existing Sapphire code base.

Scalable Graph Algorithms

Student **Paul Baginski,** University of California, Berkeley

Mentor **Van Henson,** CASC

My project involved splitting a graph into topological neighborhoods of a fixed radius. In other words, given a graph G and a natural number d, pick a subset S of the vertices such that every vertex of G is within distance d of some element of S. The vertices in S are called centers, and for a given center, the set of all vertices of G within distance d of that center is called the d-neighborhood. Our goal was to find a relatively small set of centers.

Three distinct algorithms were produced—the Degree algorithm, Tree algorithm, and Complement algorithm. We measured the effectiveness of these algorithms using the following criteria:

(a) Relative number of centers chosen
(b) Runtime
(c) Scalability of the algorithm to different values of d
(d) Amount of overlap in the d-neighborhoods
(e) Amount of information needed beforehand (e.g., adjacency matrix).

Both the Degree algorithm and the Tree algorithm performed very well under most criteria, but in each case there was a major shortcoming that made the algorithms unfit for implementation. The Complement algorithm seems most promising, but a full analysis has not yet been completed. The number of centers chosen is uncalculated, but in every other respect, the algorithm scores either well or exceptionally well.

High-Speed Multi-Component Flows

Student **Jeffrey W. Banks,** Rensselaer Polytechnic Institute

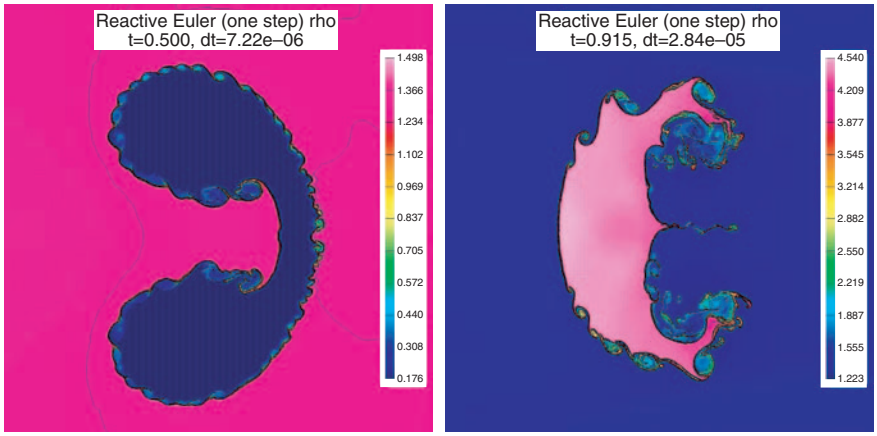
Mentor **William D. Henshaw,** CASC

We studied fluid flows with more than one component as described by the well-known multi-component Euler equations. It is a well-known phenomenon that numerical studies of these equations in conservation form can lead to unphysical pressure oscillations.

There are two main methods to combat this problem. The first is to investigate the system from a primitive variable formulation. This clearly necessitates a correction to bring the method back to a quasi-conservative form so as to maintain proper shock speeds. The other is to use some sort of conservative scheme on all but the fluid fraction variables and then advect those variables based on a special scheme that disallows the aforementioned pressure oscillations.

We present a third approach to this problem whereby one begins with a conservative method, but includes a source around contacts. This allows for all variables to be advanced through the same method while maintaining a conservative method in regions of the flow where there are no contacts between different fluids. We use an adaptive mesh refinement (AMR) scheme that is itself a constituent of an overset grid method and add the capability for moving grids. This allows our solver to operate on complicated geometries while maintaining efficiency through the AMR scheme.

We have demonstrated convergence through the method of analytic solutions and the so-called twilight-zone flows. We then performed calculations of the interaction of a shockwave with cylindrical bubbles of helium (*He*) and refrigerant-22 in order to show agreement with previous experiments and numerics. Then we used this solver to calculate the interaction of a shock with a special shaped piece of metal as in a shaped explosive charge.



On the left, interaction of shock with He bubble, and on right interaction of shock with refrigerant-22 bubble. Both pictures have the shock running left to right, are at late time after the shock has passed, and clearly show the dominance of unstable features (Kelvin-Helmholtz, Rayleigh-Taylor).

TOC

Firewall Egress Filtering

Student **Lerone Banks,** University of California, Davis

Mentor **Jerry Rayome,** CSP

During my summer internship, I worked on several projects. My primary assignment was focused on presenting a solution for implementing egress filtering on the firewalls here at LLNL. Accomplishing this task required me to gather information about current network usage and requirements. From there, I explored firewall configuration details for the Green Network. With this information, I investigated current best practices in egress filtering and presented a solution for the preliminary steps of implementing egress filtering. Some of my other duties included evaluating various intrusion detection tools, making recommendations about their suitability for deployment and suggesting ways to secure the use of the domain name system (DNS).

Volumetric Mesh Parameterization Towards Slow-Growing Subdivision

Student

Janine Bennett, University of California, Davis

Mentor

Valerio Pascucci, CASC

Scientific simulations often generate unstructured mesh domains of arbitrary topology with data values sampled at the mesh nodes. However, a hierarchical, regular mesh structure is preferred for efficient analysis and rendering of the data. Subdivision methods are one of the most successfully used techniques in multi-resolution data representation for surface meshes. Initial research has been done to extend these techniques to the multi-resolution representation of volumetric data. For example, the Slow Growing Subdivision (SGS), first described in [1], is a subdivision method that generalizes to volumetric and higher-dimensional meshes. This approach currently works well for convex domains and regular curvilinear grids, where a mapping to a regular structure is provided explicitly.

This summer, we extended SGS so that it will work on arbitrary volumetric grids with non-convex boundaries. We use a region-growing technique that starts from single-cell sets, which are expanded as long as their structural properties are maintained in order to decompose the input domain into topological balls. We have developed a two-step process to map the resulting regions to an octahedral domain, i.e., we first map the boundary vertices and then the interior mesh vertices. To map the boundary, we determine edge paths that correspond to the edges of an octahedron. Using the method proposed in [2], we map the remaining boundary vertices to the faces of the octahedron. We are currently exploring methods to map the interior mesh vertices in order to obtain the necessary isomorphism from unstructured mesh connectivity to SGS connectivity.

Publications

[1] Valerio Pascucci, "Slow growing subdivision (SGS) in any dimension: Towards removing the curse of dimensionality," in *Proceedings of Eurographics* 2002, pp. 451–460, Saarbrucken, Germany, September 2002.

[2] M. S. Floater, "Parametrization and smooth approximation of surface triangulations," *Comp. Aided Geom. Design* 14 (1997), pp. 231-250.

Adaptive Algebraic Multigrid Preconditioners in Quantum Chromodynamics

Student **James Brannick,** University of Colorado at Boulder

Mentor **Rob Falgout,** CASC

Standard algebraic multigrid methods assume explicit knowledge of so-called *algebraically smooth* or *near-kernel* components, which, loosely speaking, are potentially large errors that correspond to relatively small residuals.

Typically, these methods automatically generate a sequence of coarse problems under the assumption that the near-kernel is locally constant. The difficulty in applying algebraic multigrid to lattice Quantum Chromodynamics (QCD) is that the near-kernel components can be far from constant, often exhibiting little or no apparent smoothness. In fact, the local character of these components appears to be random, depending on the randomness of the so-called "gauge" group. Hence, no *a priori* knowledge of the local character of the near-kernel is readily available.

Our work develops an adaptive algebraic multigrid (AMG) preconditioner suitable for the linear systems arising in lattice QCD. The method is a recently developed extension of smoothed aggregation, aSA, of Brezina, Falgout, MacLachlan, Manteuffel, McCormick, and Ruge, in which good convergence properties are achieved in situations where explicit knowledge of the near-kernel components may not be available. This extension is accomplished using the method itself to determine near-kernel components automatically by applying it carefully to the homogeneous matrix equation, $Ax=0$. The coarsening process is modified to use and improve the computed components. Preliminary results with model 2D QCD problems suggest that this approach may yield optimal multigrid-like performance that is uniform in matrix dimension and gauge-group randomness.

A Memory Insensitive Format for Out-of-Core Access to Unstructured Volumetric Meshes

Student **Steven Callahan,** University of Utah

Mentor **Valerio Pascucci,** CASC

Unstructured volumetric meshes are the preferred data format for scientific simulations. With the size and complexity of these meshes increasing every year, techniques have been developed to visualize the pertinent information quickly and efficiently. This visualization is a commonly performed preprocessing step in which a triangulated isosurface is extracted for a particular isovalue. For large meshes, this extraction is performed out of core. The work I have done introduces a global indexing scheme that accelerates the traversal of the data structure on disk by improving the spatial locality of the data. In addition, the data is converted to a coherent streaming format that facilitates IO-efficient, out-of-core algorithms.

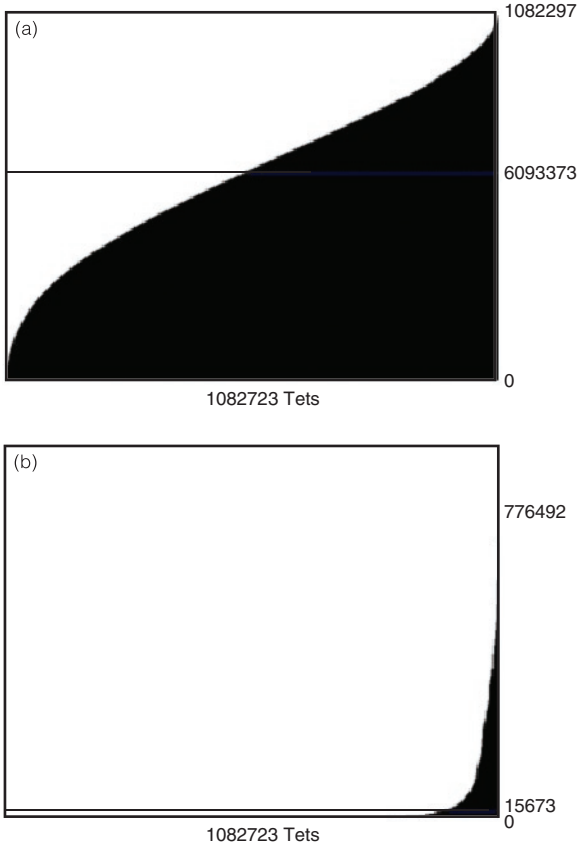


Figure 1. Comparison of original vertex order (a) and Lebesgue order (b)

The key idea is to order the vertices in such a way that minimizes the external memory accessing needed to traverse the volumetric cells. An out-of-core technique was developed to re-index the vertices using a Lebesgue space-filling curve and output the data in a streaming format. This new global indexing was then compared with previously proposed techniques using tools that displayed the IO cost of volumetric cell traversal. The contributions of my work include an out-of-core algorithm for re-indexing arbitrary unstructured volumetric meshes, an extension of streaming meshes for unstructured volumetric data, and a technique for analyzing the efficiency of the resulting mesh.

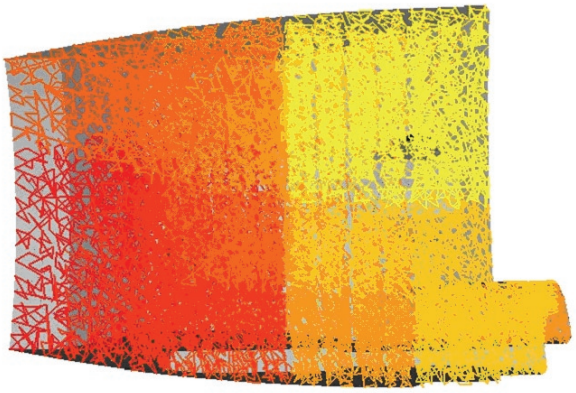


Figure 2. Lebesgue order imposed on the vertices of an unstructured mesh.

TOC

Discovery Center Display

Student **Jedidiah Chow,** Granada High School

Mentor **Jean Shuler,** CSG

The goal of our project was to produce an interactive program that visitors could use in order to inform them of the computing goals and accomplishments of the LLNL Computational Directorate.

To keep things simple and editable, only well-known programs were used to make the project. Microsoft Power Point is the basis for the display program, and it proved adequate for navigating through various topics, as well as being aesthetically pleasing. Basically, the program elaborates on supercomputing history at LLNL, visual display equipment, and the recent uses of the supercomputers. The project was completed under the supervision of Terry Girill and Bob Howe.

As with any program, periodical general maintenance and modifications will be necessary, and major updates will be accessed as time progresses and the information presented becomes out of date.

Incorporating Electrokinetic Effects into the EB Navier–Stokes Embedded Boundary Incompressible Fluid Solver

Student

Kevin Chu, Massachusetts Institute of Technology

Mentor

David Trebotich, CASC

Motivated by the recent interest in using electrokinetic effects within microfluidic devices [1], we have extended the EB Navier-Stokes embedded boundary incompressible fluid solver [2] to be able to handle electrokinetic effects. With this added functionality, the code will become more useful for understanding and designing microfluidic devices that take advantage of these effects (e.g., pumping and mixing).

Supporting the simulation of electrokinetic effects required three main extensions to the existing code:

(1) Addition of an electric field solver
(2) Development of a module for accurately computing the Smulochowski slip-velocity at fluid-solid boundaries
(3) Extension of the fluid solver to handle non-uniform, inhomogeneous Dirichlet boundary conditions.

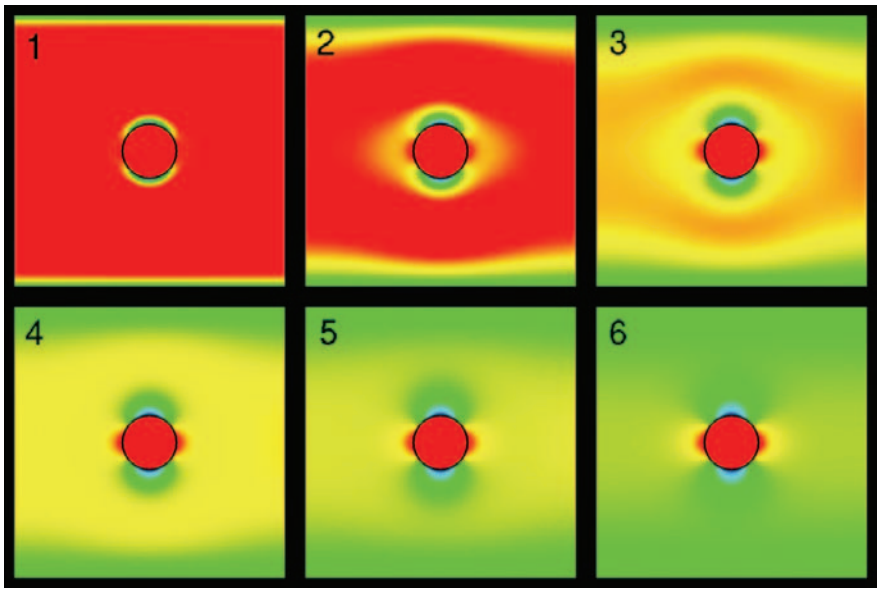
The first and second extensions were needed to compute the electrokinetically generated slip-velocity at fluid-solid boundaries. The third extension made it possible for the fluid flow to be driven by a slip-velocity boundary condition rather than by a pressure difference between inflow and outflow.

We are in the process of verifying and validating the code. Figure 1 shows the time evolution of the horizontal component of the fluid flow around a single cylindrical post where the flow is driven by electrokinetic slip at the surface of the posts and the top and bottom walls. In the future, we hope to use the code to simulate electrokinetically driven flows in novel microfluidic devices being explored for biological applications.

References

(1) T. M. Squires and M. Z. Bazant, "Induced-charge electro-osmosis," *J. Fluid Mech.*, 509 (2004), pp.~217–252.

(2) D. Trebotich, "Working Notes for Higher-Order Projection in Embedded Boundary Framework," private communication.



A flow driven by electrokinetic slip moves around a cylindrical post

Applying Morse Theory to Computational Data Sets

Student

Kree Cole-McLaughlin, University of Utah

Mentors

Valerio Pascucci, CASC **& Terence Critchlow,** CASC

Morse theory is a powerful tool that connects the theory of real-valued functions over a manifold to the topology of the manifold. Recently, computer scientists have begun to develop computational tools for generating topological characterizations of scientific data using Morse theory. The contour tree, for example, characterizes how the connected components of the level sets evolve as the function value increases.

Another important characterization is the Morse-Smale complex, which decomposes the data set into regions of uniform gradient flow. This is a more complete characterization than the contour tree and is in some sense the essential application of Morse theory. Algorithms for computing the Morse-Smale complex over the 2- and 3-spheres have been developed within the last few years.

The major problem in applying Morse theory to computational data sets is that the theory was developed for smooth manifolds and differentiable functions. Previous techniques simulate the differentiability of the function and use combinatorial data structures to keep track of several types of degeneracies that can arise. In two dimensions this does not cause much of a problem, however, for three-dimensional data sets, it adds a great deal of complexity to the description and implementation of the algorithm.

Recently Prof. Robin Forman developed a discrete version of Morse theory that can be thought of as generalizing the smooth theory. Forman's version of the theory has faced some problems in finding applications, due mainly to the fact that functions are defined in an apparently unnatural way. However, over the summer, we developed a technique for applying Forman's theory in such a way that we can prove that the results are equivalent to the previous techniques.

The benefits of our approach are many. Most significantly is the simplicity of the algorithms, which require no additional data structures for dealing with degeneracies. In some sense, we push all the complexities of the previous approaches into the theoretical framework and deal with them in the proofs of theorem. Our algorithms, on the other hand, remain simple. Additionally, our approach has the potential for being extended beyond three-dimensional data sets, where Morse theory can play a vital role in understanding the data, since visualization is nearly impossible. Finally, we do not need as many restrictions on the underlying manifold as the previous techniques, which only work on spheres. This technique is still being worked out, but our work this summer has proven that it is valid.

TOC

Geometric Multigrid for Variable Coefficient Maxwell's Equations

Student **Dylan Copeland,** Texas A&M University

Mentor **Barry Lee,** CASC

We are interested in numerically solving Maxwell's equations of electromagnetics with variable coefficients. Variable coefficients are difficult because the bilinear form that corresponds to the variational formulation has a large near-nullspace consisting of more than just gradients. Consequently, methods such as Hiptmair's hybrid smoother are ineffective on some components of the error in the near nullspace.

We implemented a multigrid method that uses interpolation operators designed so that near-nullspace components are locally interpolated from the coarse grids. To approximate curl-free components of the solution, we solved the corresponding scalar Poisson equation with the Maxwell equation simultaneously. Interpolation operators were then needed for scalar and vector finite element spaces. For the scalar spaces, we used BoxMG interpolation, which is known to be very efficient for Poisson equations on structured grids. The interpolation for the Nedelec vector finite element space was constructed locally so that coarse grid functions were interpolated to fine grid functions in the near-nullspace of the operator associated with the variational formulation. This was accomplished by the inversion of local element matrices, which is quite inexpensive.

To test the efficiency of our multigrid method, we performed numerical experiments with coefficients containing oscillations and jumps. The number of iterations required by the V(1,1) cycle to reduce the L2 norm of the residual by eight orders of magnitude was approximately 40% less than that of Hiptmair's multigrid method. Thus, a significant improvement in the convergence rate is achieved with our method.

TOC

Web Designer, Program for Climate Model Diagnosis and Intercomparison

Student **Steven Davis,** California State University, Hayward

Mentor **Charles Doutriaux,** EEBI Division

My assignment was to make the previous version of the PCMDI Web site http://www-pcmdi.llnl.gov/old easier to use and more aesthetically pleasing to the user. This was accomplished by creating a consistent design for the entire site, recreating the directory structure to make it more understandable, and implementing new features that allow users to navigate more easily, such as a drop-down menu system, breadcrumb navigation, and a search box on every page.

These features, and many smaller details, were added and changed to convert the PCMDI Web site to what is today http://www-pcmdi.llnl.gov. I will continue to work one day a week during the school year to add more functionality to the site, such as database-driven dynamic pages, as well as the ability for group members to update certain pages without compromising the design.

Edge Betweenness Properties in Complex Networks

Student

Stanko Dimitrov, University of Arizona

Mentor

Edmond Chow, CASC

Vertex betweenness, the number of shortest paths that pass through a vertex, has been studied extensively in social networks; the notion of vertex betweenness was later extended to include edge betweenness. These measures can be thought of as ranking the vertices (in the case of vertex centrality) and the edges (in the case of edge centrality) in order of importance. This importance can help people to prevent epidemics from spreading by removing people with the highest rank, or seeing that one edge is a bridge between communities (a cluster of vertices). However, no formal, mathematically formulated definition and exact algorithm for edge betweenness was presented.

We initially formally defined edge betweenness and presented a fast algorithm for computing edge betweenness by modifying the current fast algorithm for computing vertex betweenness. With our formal definition of edge betweenness, we investigated how edge betweenness behaves in different types of complex networks, where a complex network is a large graph with a specific probability distribution on its vertices that is used to determine the edges of a graph. The vertex probabilities help determine which two vertices make up an edge, meaning the vertex with the highest probability is more likely to make up an edge than a vertex with the lowest probability.

In our experiments, we had three types of complex networks—random, spatial, and scale free. We also had one real network that was part of the network of routers that make up the Internet. We observed how the measure behaves in our models and how it behaves in the real network. We need to know how our measure behaves so we can classify our models. With this classification, we hope to be able to say something about the structure of a real network given that we know how our measure behaves in that network.

One drawback to edge centrality is it takes a long time to compute, which means we are interested in finding other measures that can approximate edge centrality. With these approximations, we can quickly rank the edges and begin making statements about the structure of our complex networks.

TOC

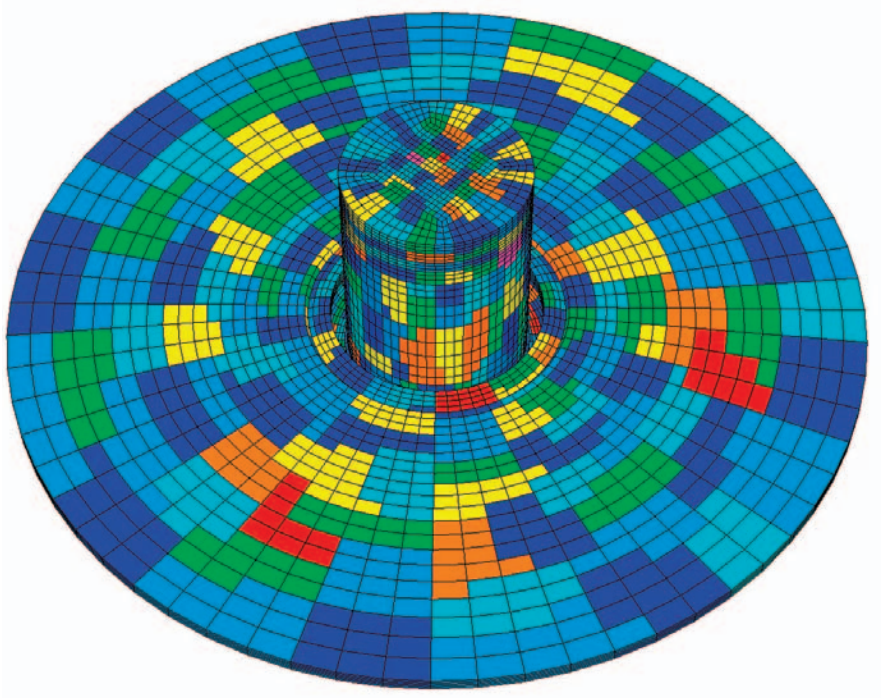
Element Agglomeration AMGe Solvers for Unstructured Finite Element Problems

Student **Veselin Dobrev,** Texas A&M University

Mentor **Panayot Vassilevski,** CASC

Unlike the standard (geometric) multigrid algorithms, which require a sequence of coarse operators and interpolation/restriction operators, the AMG methods only require a fine-grid matrix as input, which is used to construct the coarse level matrices. Similarly, the element agglomeration AMG (AMGe) methods, which can be used when the finite-element method (FEM) is used for discretization, also use fine-grid information only—the element matrices combined with topological information for the FE mesh and discretization space. Therefore, these methods can be used for problems on unstructured meshes and complex domains where standard multigrid is not applicable.

Based on the Texas A&M finite element package AggieFEM, a code was developed for discretizing various types of PDEs in 2D and 3D, including second-order elliptic equations, linear elasticity, etc. The code uses arbitrary triangular (2D), tetrahedral and hexahedral (3D) meshes, allows for global (and in some cases local) refinement, and supports several types of finite elements—linear and quadratic conforming, and linear nonconforming. The generated discretizations were used as input to an AMGe library written by Dr. Vassilevski for testing and experimental purposes and to evaluate the solvers' performance. The developed code also provides visualization for the obtained solutions and element agglomerations (generated by the AMGe code). As a result, improvements and enhancements were made to the library. In particular, a new agglomeration algorithm was developed based on the graph partitioning library METIS.



Element agglomeration in 3D

TOC

Parallel Implementation of a FETI–DP Method for Elliptic Problems with Mortar Finite Element Discretization

Student

Nina Dokeva, University of Southern California

Mentor

Panayot Vassilevski, CASC

This project focused on elliptic problems with discontinuous coefficients discretized by Finite Element Method (FEM) on non-matching triangulations across the interface using the mortar technique. The resulting problem is solved by a Finite Element Tearing and Interconnecting Dual Primal (FETI-DP) method proposed by M. Dryja and O. Widlund, and generalized by adding multiple scalings in the preconditioner to problems with discontinuous coefficients by N. Dokeva, M. Dryja and W. Proskurowski. The convergence has polylogarithmic dependence of the mesh-sizes. Moreover, it is independent of the number of subdomains which makes the algorithm suitable for parallel implementation with one processor working on each subdomain. The parallel code was developed using C and MPI. It ran successfully on 256 processors.

After domain decomposition, the problem is divided into three types of tasks:

(1) Solvers on the subdomains (with different meshes of discretization) which run individually and in parallel

(2) A problem on the interfaces between the subdomains which can be solved in parallel with only a few global communications

(3) A problem on the cross-points between the subdomains, which is a global task

Scalability when increasing the number of processors was only acquired after optimizing the individual parts of the algorithm, paying the most attention to the problem with the Schur complement matrix on the cross-points between the subdomains. Ultimately, the algorithm proved to be scalable when increasing the number of processors, with only a minimal overhead in communications, which significantly decreased the time for executing the program compared to the sequential algorithm.

Future work will aim at improving the cost increase when refining the mesh, making further improvements in the preconditioner, and considering different mortar functions.

TOC

Graph Viewer Improvements

Student **Emily Eder,** UCLA

Mentor **Robert Fernandes,** IOAC

My assignment was to make improvements to the Java-based graph viewer currently in use at the Lab. The first part of my assignment was to make a URL Mapping Panel. This was accomplished by creating a Java dialog box to input corresponding Web locations and file locations and addressing the need to store and retrieve corresponding Web locations and file locations.

The second part of my assignment was to conduct research into graph layout algorithms for the graph viewer. My research focused on force-directed algorithms and Java implementations of these algorithms, including the Geometric Efficient Matching (GEM) algorithm, the spring embedder algorithm, the simulated annealing algorithm, and graph drawing by stress majorization. The goal of the project was to identify the most efficient force-directed algorithm, which would then be implemented in the graph viewer within the next six months. The GEM algorithm was identified as one of the most effective force-directed algorithm.

Platform Independence for CALE: wxWidgets and GNU Autotools

Student

Christopher Egner, Rochester Institute of Technology

Mentor

Paul Amala, A Division

My work centered around two objectives, implementing a graphical user interface for C-language Arbitrary Lagrange Eulerian (CALE) using wxWidgets and reimplementing CALE'S configuration and build system using GNU Autotools. CALE'S user interface was based in X11 using a single-threaded architecture. As such, there were no concurrency issues, and the X11 API calls, while wrapped by other functions, were more coupled to the physics code than would be standard in a modern Model-View-Controller design.

In order to combine the wxWidgets and CALE architectures, a multi-threaded model was chosen last summer. This introduced two related problems. First, CALE'S code was not designed to be called in more than one thread, so it is not well-suited to a normal multi-thread model. Second, wxWidgets is not thread-safe, imposes the requirement that all graphics calls (an ill-defined term) be made from the main thread, and assumes that the program's execution architecture can be governed by its event-processing loop. It is these architectural issues that are at the foundation of many of the problems I encountered, and some remain unresolved. Completed work includes an implementation of buffered graphics, improved efficiency, increased stability, patches to wxWidgets, and a change in template models for wider compiler compatibility.

CALE'S former build system was based on an M4 script that produced a single makefile to build either the high-explosive or inertial-confinement fusion version of the code with manual platform detection and manual dependency tracking. Reimplementing the configuration and build system with GNU Autotools resulted in an automated and more customizable configuration system, as well as a more maintainable, faster build system with automated dependency tracking. Combined, the new systems configure and build not only the high-explosive and inertial-confinement fusion versions of CALE, but also the project's supporting tools and documentation.

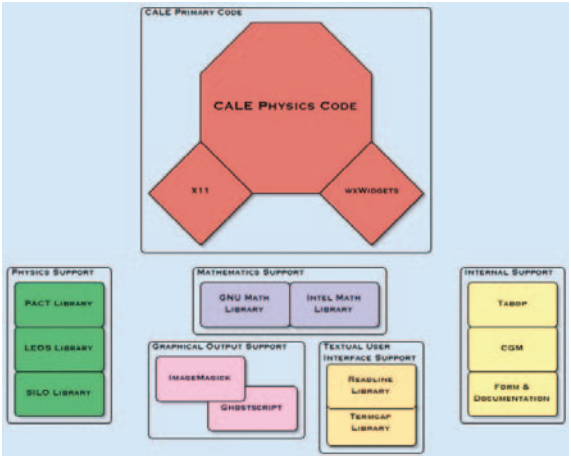


Figure 1. The structure of CALE and its supporting libraries and tools, grouped by purpose. CALE can be built with either the X11 or wxWidgets GUI and with any supporting libraries available.

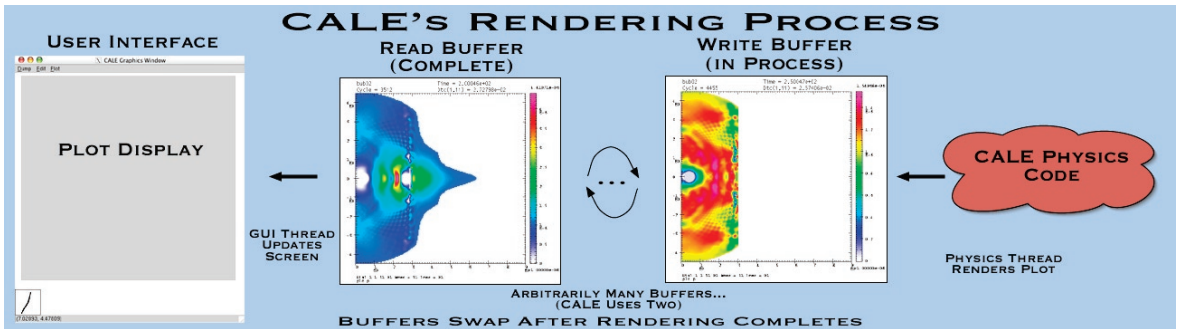


Figure2. CALE'S buffered graphics rendering system. CALE can be used to render animations as it computes. However, animation can lead to flickering on the screen as the program redraws. Thread-safe, rotating buffers assure that CALE can always refresh the screen, even as it works on the next plot. By storing the last rendering, CALE can load it quickly to the screen in order to repair damaged areas of the screen without having to recompute. By computing the entire plot first, then drawing a bitmap rapidly to the screen, flicker is appreciably reduced and drawing is significantly faster.

TOC

Wavelet-Based Opacity Data Compression

Student **Christopher Elofson,** University of Arizona

Mentor **Vijay Sonnad,** DCOM

Astrophysical opacity data sets are large and often contain many spectral lines. A compression algorithm capable of accurately representing both the peaks and depths between them (related to a quantity known as the Rosseland mean) is desirable, as this will provide greater computational efficiency.

While compression could be implemented using a Fourier transform, the Fourier method cannot be inexpensively adapted for data with many peaks, as the total number of samples must be increased to improve local resolution.

A better approach to opacity data compression is to use Discrete Wavelet Transforms, which differ from Fourier transforms in that they can use any function as a basis function—Fourier can use only sinusoidal functions. Wavelets are not only a better match to data's end behavior, they are also a much less expensive, yet equally accurate, representation of spectral lines, made possible by wavelets' capability of local resolution adjustment.

Through the testing of six varying opacity data sets, the Discrete Wavelet Transform proved to be a very effective way to increase manageability of opacity data. In five of six cases, the linear transform provided the best results, with compression rates varying from 45–99% (three of six cases yielding rates over 98%) while conserving the Rosseland mean within 0.00001%—the Haar basis varied between 38–96% compression. In the case where many spectral lines were densely packed over the entire domain, the Haar basis performed better in data representation, producing a lower rate of error in the Rosseland mean as well as lower absolute error in comparison to the linear compression.

TOC

ODE Visualization and Archive Tool in Python

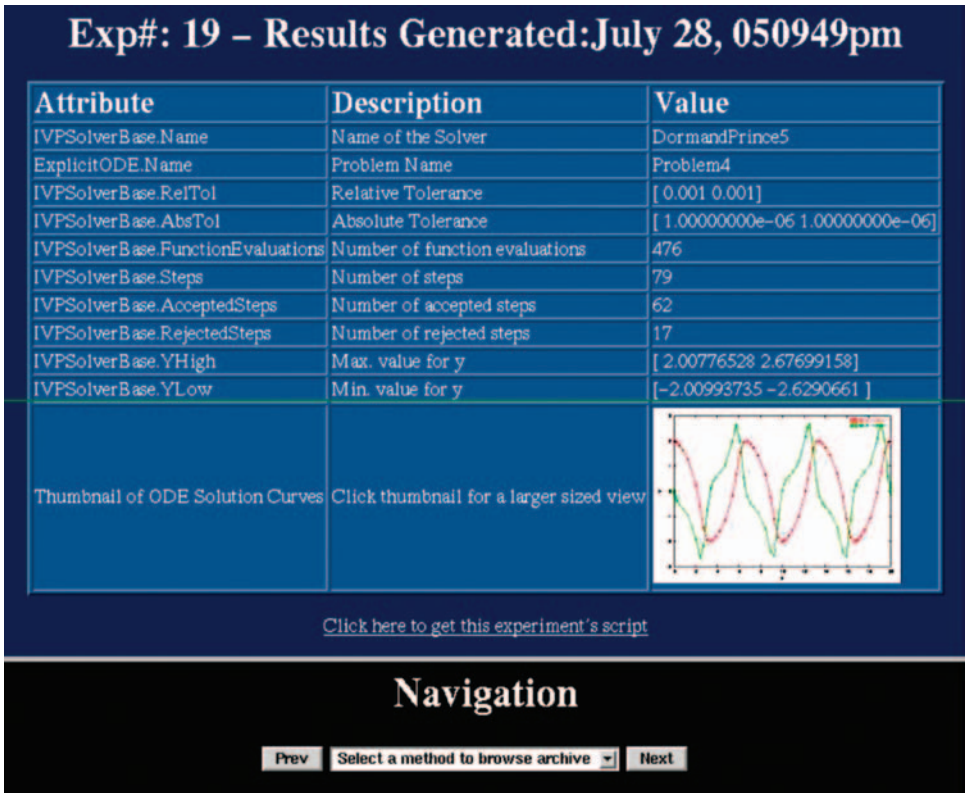
Student

Dayo Esho, University of California, Berkeley

Mentor

Steven Lee, CASC

The Runge-Kutta method is a numerical technique used to approximate the solution of ordinary differential equation (ODE) initial-value problems. We developed a visualization and archiving tool to interface between the data output of advanced ODE solvers. In particular, we studied higher-order, adaptive-stepsize, explicit Runge-Kutta methods by Dormand and Prince (D&P) and Calvo, Montijano, and Randez (CMR). By analyzing the visual output from these solvers through our tool, we can compare solutions between solvers, determine their relative accuracy and efficiency, and monitor solver performance statistics. Archiving these results in a Web interface allows for rapid regeneration of particular experiments and easy modification of parameters to extend an analysis of a particular solver-problem combination. Our work focused on the ability of these solvers to handle stiff ODEs.



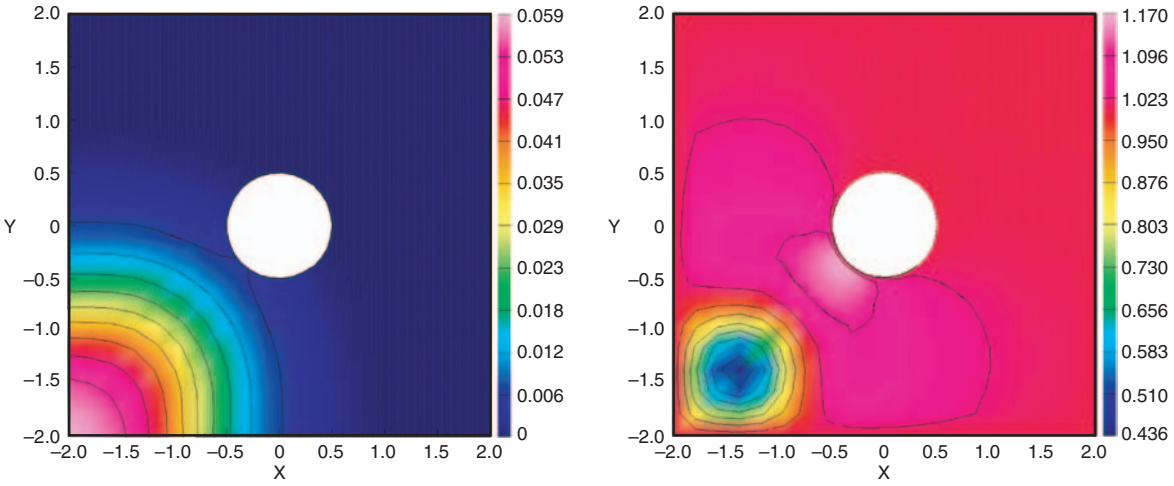
Screenshot of the ODE Viz/Archive Tool

2D Numerical Modeling of Soap Film on Overture

Student **Pak-Wing Fok,** Massachusetts Institute of Technology

Mentor **Petri Fast,** CASC

My assignment was to learn how to use Overture, a set of C++ Libraries that facilitate the solution of partial differential equations in complex geometry, and then use that knowledge to code up a solver for a set of soap film equations recently derived by Dr. Petri Fast. Following this, we simulated the short-time evolution of a soap film, in the middle of which a solid cylinder was immersed (see below). I also performed some error analysis and convergence studies to validate my numerical solution.



Two snapshots from my numerical runs. The one on the left shows surfactant concentration; the other shows the thickness of the film.

TOC

Implementing Cray Pointers in the GNU gfortran Compiler

Student

Camille Fournier, University of Wisconsin

Mentor

Mike Kumbera, DCOM

During my assignment, I began the implementation of Cray pointer support in the GNU gfortran compiler. Specifically, I modified the parser to recognize Cray pointer statements, and modified the gfortran frontend to produce code that correctly represented these statements to the gcc backend so that it could compile them. This process included creating a new data type for Cray pointer data, adding command-line arguments to specify whether these pointers were used in the program, and adding intrinsic functions that were required in order to use pointers of this type.

Performance Analysis of Monitoring and Information Systems Using NetLogger and an Investigation into P2P Overlay Topologies

Student

Jeffrey Freschl, University of Wisconsin, Madison

Mentor

John Johnson, DCOM

In collaboration with members of Globus at Argonne National Laboratory (ANL), we analyzed and compared the performance of three monitoring and information services—the Relational Grid Monitoring Architecture used in the European Grid test bed, MDS2 used in the Globus toolkit, and Hawkeye used in Condor pools. Our analysis involved instrumenting the source code with NetLogger calls to locate performance bottlenecks. With the instrumented tools, we simulated up to 600 clients, 600 monitored machines, and 80 sensors. Our results indicate that running the main components of such systems at well-connected sites (i.e., the network is a bottleneck) and caching the monitoring information (i.e., activating an information sensor is a bottleneck) can provide significant performance improvements. In addition to completing a poster presentation, we are preparing a journal paper.

My work on peer-to-peer overlays involved a survey of the current work on the overlay mismatch problem. The main methods include using landmarks, clustering to partition the overlay, and locally optimizing a peer's connections to its neighbors. The main use of current P2P networks involves file sharing (e.g. Kazaa), and the overlay mismatch does not pose much of a performance problem. However, someday we may be able to utilize the Internet for far more compute-intensive, parallel algorithms (not necessarily restricted to embarrassingly parallel algorithms) where communication can be a significant bottleneck. Would it be possible to use a P2P network for tightly coupled simulations? Perhaps not to that degree, but high-throughput computing could utilize a P2P network as a computational resource.

Bringing Order to CHAOS

Student **Evan Geller,** Summit Preparatory High School

Mentor **Albert Chu,** HPSD

My major project focused on Cerebro, a host monitoring system. I coded the networking piece and config file parsing that enabled system administrators to work without having to make a config file for each monitored node.

I also created reproducers and integrated them into our test suite to chronicle the following issues:

- A bug found in the cpickle function in Python which caused a memory leak
- A bug found in the Network File System (NFS) that resulted in chmods inconsistency across nodes
- A bug, found in the getrusage system call that allowed system time to be reported incorrectly
- A bug that was found in the interaction between rsync and NFS that caused stale file handles.

In addition to this work, my mentor found a test to assess various aspects of NFS, which I was able to integrate into our test suite and report on NFS reactions. Our group needed a script that would monitor how NFS reacted to various commands. To address this need, I wrote a script that wrapped around the given command and observed how NFS reacted.

Assessing Performance of Hybrid MPI/OpenMP Programs on SMP Clusters

Student

Tobias Gradl, Technical University of Munich

Mentor

Edmond Chow, CASC

The SMP cluster architecture is the prevalent type of computer system in today's high performance computing landscape. Two strategies are popular for programming these computers. In the *pure MPI approach,* message passing is used for all data exchange, even between processors within a node. In the *hybrid approach,* messages are only passed between processors on different nodes, while processors within a node share data through multithreading.

While hybrid programs tend to demand more coding effort, they promise to be faster than pure MPI programs. Practical experiments show, however, that in many cases both strategies are equally fast; some even show that pure MPI is faster. Some of the reasons for this behavior are well understood, while others are not. The insights gained from the comparison will help explain the behavior of large real-world applications. They can also guide programmers in the design of new applications.

We used a simple but realistic benchmark, the matrix-vector product, to compare the two programming strategies. The simplicity of the matrix-vector product helped us analyze and understand its behavior in detail.

We ran the benchmark program using different problem sizes and numbers of processors and took timings of all code sections. The characteristics of the collected timings were then explained by various environmental factors like cache size, memory bandwidth, network bandwidth, etc. Knowing how these factors affect the program performance enables us to extrapolate relevancy to other applications and system configurations.

TOC

Censorship Resistance

Student **Rachel Greenstadt,** Harvard University

Mentor **Terry Brugger,** NAIC

I participated in three main projects during the Lab's summer program: Circumventing Government Firewalls, LOCKSS (Lots of Copies Keep Stuff Safe), and Indexing Hacker Tools.

Many countries around the world, such as China, Singapore, Iran, and Saudi Arabia, have censored access to the Internet. Traditionally, these firewalls have been circumvented with proxy servers, which are also eventually censored. This project explored efforts to win this arms race and identified key bottlenecks where more work is needed. We concluded that circumvention of such systems is easy to accomplish on a small scale, but difficult on a large scale. The hardest problem seems to be publicizing information about the circumvention system without having the system blocked.

LOCKSS is a peer-to-peer system that allows digital libraries to collaborate and archive their data. It aims to protect information stored on a single system from being distorted due to bit rot, targeted censorship, or physical destruction. The LOCKSS project includes participants at multiple universities and a deployed beta system with 80 participating libraries. Libraries periodically poll a sample of peers in the system to determine if their copy of the document is good. Pollees consist of a combination of trusted "friends" and other discovered peers. In the event there is no clear supermajority of peers, an alarm is called as an intrusion detection measure. We conducted simulations of a network of 1,000 peers wherein each peer conducted 3,600 polls to determine the most effective, practical mechanism for a peer to respond to alarms. This mechanism proved to be a procedure in which peers who experienced an alarm contacted and healed their friends.

For the third project, I anonymously surfed a variety of hacker and security Web sites in order to create an HTML table describing the range of hacker tools currently available in the open source community and indexed them by category.

TOC

Hierarchical Morse–Smale Complexes

Student **Attila Gyulassy,** University of California, Davis

Mentor **Valerio Pascucci,** CASC

A Morse–Smale complex is a structure that represents the topology of a data set. Each cell of the complex consists of a minimum, a set of 1-saddles and 2-saddles, and a maximum. Connectivity of these critical points is also stored in the cell. The Morse–Smale complex can be simplified for use in applications such as topological smoothing or multi-resolution viewing.

We developed an algorithm for simplifying Morse–Smale complexes that relies solely on combinatorial decisions, thereby avoiding numerical instability. Simplification of a Morse–Smale complex is possible through the cancellation of critical point pairs. There are two types of cancellations—saddle–extrema cancellations and 1-saddle–2-saddle cancellations. We developed rules for determining when a cancellation is valid in a complex, and for reconnecting the simplified complex.

A multi-resolution representation of a Morse–Smale complex can be achieved through a simplification hierarchy. We presented a set of rules that govern valid simplifications of a complex and developed a hierarchy based on the independence of those cancellations. We declare two cancellations to be independent when their affected areas in the Morse–Smale complex do not overlap. We represent the hierarchy as a directed acyclic graph that encodes the independence of cancellations. A multi-resolution reconstruction of the complex can be attained by cutting the hierarchy graph.

Future work involves extracting a Morse–Smale complex from a scalar data set and applying the simplification hierarchy to topology-based smoothing of actual data sets.

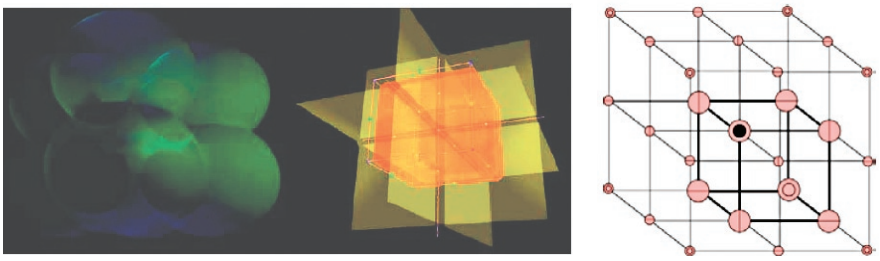


Figure 1. An isosurface of a simple data set (left), the corresponding stable and unstable manifolds, and the associated Morse-Smale complex.

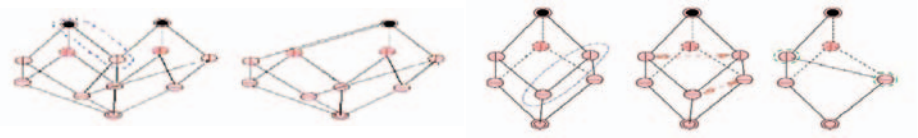


Figure 2. A saddle-maximum cancellation (left) and a 1-saddle–2-saddle cancellation.

MPI Profiling

Student **Daniel Han,** University of Southern California

Mentor **Terry Jones,** Services & Development Division

Message Passing Interface (MPI) is the de facto standard in writing massively parallel programs and has been embraced by the scientific community for use in solving a variety of problems. Performance is often a crucial factor, especially of grand challenge problems. Although there have been many studies on the scalability of these applications, there have not been many that focused on the specific types of MPI calls being made and their relative importance. Using a profiling tool called mpiP, I surveyed a large spectrum of parallel scientific applications and analyzed the results.

Parallel Analysis of Asymmetries in Symmetric Simulations in ALE3D

Student **Kevin Hoffman,** Brigham Young University

Mentor **Robert Cooper,** DCOM

The unstructured finite-element mesh hydrodynamics simulation code ALE3D efficiently scales to problems with millions of zones. Validation of the correctness of the code as it is developed is critical to maintaining the accuracy of the simulation. One way to validate the code is to simulate a symmetrical problem and ensure that a reasonable level of symmetry is maintained in the solution fields.

I designed and developed a parallel analysis tool to determine the amount of asymmetry at each zone and node in the mesh. Some challenges included efficiently determining communication needs between processors, efficiently pairing each point in the mesh with its symmetrical counterpart, and dealing with vector-based variables.

I further enhanced the algorithm to support taking the difference between two fundamentally different meshes, such as problems meshed at different resolutions or a mesh partitioned using different methods. This allowed me to verify the preservation of the order of operations using different domain partitionings.

I also developed a parallel statistics calculation framework that calculates various statistics over different subsets of the mesh—region, domain, problem, etc. The framework allows for easy addition of new statistics.

Finally, I researched how to redesign the domain partitioning code to meet the future needs of the ALE3D team. I created a flexible design that abstracts the interface between the mesh and the partitioner, allowing for various mesh sources, chaining of partitioners, refinement of partitioning, and dynamic repartitioning. I also wrote a partial implementation showing the validity of the design.

Combinatorial Feature Extraction for a Streaming Framework

Student

Taylor Holliday, University of California, Davis

Mentor

Valerio Pascucci, CASC

We devised and implemented combinatorial algorithms for constructing discrete Morse functions and Morse complexes on 2D and 3D simplicial complexes. Based on spanning trees, this combinatorial approach is easy to implement, amenable to a streaming multiresolution representation of the data, avoids the numerical instability of previous approaches, and scales to large data sets.

Morse complexes are useful for multiresolution visualization, topological smoothing, and mesh reparameterization. They simplify a data set while preserving extrema.

Our approach builds a discrete Morse function from a piecewise-linear function on the vertices. The algorithm partitions the data set into subcomplexes, which are ordered by function value, and builds minimal Morse functions on each subcomplex. Together, these minimal functions form the Morse function for the entire complex.

Future work includes adapting these algorithms to an out-of-core hierarchical representation of the data set, such as the Z-order space-filling curve.

The figure illustrates the 2D algorithm run on a trivial simplicial complex. A minimum (yellow), a maximum (magenta), and six saddles (cyan) are depicted as colored squares. The magenta and yellow lines indicate paths of steepest ascent and decent (respectively) from the saddles. These lines are the 1-cells (edges) of the Morse complex.

TOC

Virtual PCR (vPCR)

Student

Gary Hon, University of California, San Diego

Mentor

David Clague, EETD

I worked with University of Virginia summer student John Frankovich to integrate the nonlinear equation solver KINSOL into vPCR. The overall goal was for vPCR to solve a system of free energy and conservation equations that were both highly coupled and nonlinear. This involved deriving and experimenting with half a dozen different approaches. The final approach involved improving KINSOL to support arbitrary precision arithmetic using Lawrence Berkeley National Laboratory's ARPREC package and incorporating a free energy error term.

I improved vPCR by making simulations of higher-order reactions possible for both kinetics and thermodynamics codes. The original code only supported bimolecular reactions where each reactant was a monomer. The extended code does not have this limitation. This improvement involved completely restructuring the underlying data structures and the higher-order codes that use them.

Furthermore, I added the ability to simulate polymerase extension of annealed monomers. This purely kinetic simulation involved finding extension products and then running Gillespie's stochastic chemical kinetics to form products.

Lastly, I tested and verified the functionality of all of the aforementioned modifications.

A Two-Grid Method for Radiation Diffusion

Student **Jason Howell,** Clemson University

Mentor **Carol Woodward,** CASC

Applications of solvers to nonlinear radiation diffusion problems require repeated solutions of a linearized version of the problem, and these linear problems are often the bottleneck for the entire solution process. Our approach is a two-grid scheme in which the problem is solved very accurately on a coarse mesh followed by a single solver of a linearized version on the original problem mesh. The motivation for the method is that the coarse-mesh solution will capture the nonlinear behavior of the problem, and the fine grid solver will refine the solution on a precise scale, resulting in a faster overall solution process. This method was successfully applied to a similar problem arising in ground water applications by Wu and Allen in 1999. The main theoretical development for this method comes from Xu (*SIAM Journal of Numerical Analysis*, 1996), and the extensions to our nonlinear system come from Dawson, Wheeler, and Woodward (*SIAM Journal of Numerical Analysis*, 1998).

This method has been implemented at Lawrence Livermore National Laboratory using PVODE as the ODE time integrator and transport3d as the radiation diffusion model driver. Investigations of implementation issues addressed this summer included the transfer of solution data between the coarse and fine grids, and proper application of boundary conditions to the corrections solved for each grid. Many numerical experiments were run on linear and nonlinear problems, and the performance of the two-grid solver was compared to that of a standard nonlinear solver. Different choices of physical parameters provided a range of nonlinearity for our tests.

TOC

DHS Countermeasures Test Bed – Command Center

Student **David Hoyt,** Brigham Young University

Mentor **James Schek,** NAIC Division

The initial goal of this summer project was to design an online interface to the advanced radiation detector's embedded HTTP server. This included working on the layout, design, image editing, JavaScript programming, and flow. The design was presented to the end users, who made suggestions for minor changes, and it was unanimously accepted.

Thereafter, threading and Win32 API errors necessitated alterations to the existing serial port interface code. I was able to successfully implement a testing program and provide a new solution to the problem by rewriting major portions of the existing classes. I conducted research and testing to understand important programming concepts, such as threading and asynchronous input and output.

I then began to implement a security model in Java for the Command Center utilizing Java Data Objects (JDO) and a Pointbase database. As part of my work, I had to become familiar with JDO, including its implementation, metadata files, and integration with relational databases. I created a custom tag library for use in Java Server Pages (JSP) to complement the security model. It was then necessary to unify and centralize the logins.

PerfTrack

Student

Kevin Huck, University of Oregon

Mentor

Brian Miller, CASC

Our goal was to create a tool to help scientific programmers answer difficult questions about application performance, given that the source code, build parameters, runtime environment, and hardware vary over time. PerfTrack was developed to explore technologies in parallel performance measurement, modeling, analysis, and prediction. Performance data and the associated environment data is being stored in a relational database. This database provides a foundation to build analysis tools that are scalable to large numbers of threads (over 1,024) and that are capable of comparing multiple executions. The tools will be automated to gather, store, and analyze data, in order to encourage their use in the software development cycle.

In addition, I researched the use of data mining (cluster analysis) techniques on performance data by applying techniques used in phase analysis of applications to parallel performance data. As a result of this research, some initial comparisons between threads in a single execution were made. The results of the analysis were similar to other studies using cluster/factor analysis on the same application.

A Multilevel Preconditioner for a PDE-Constraint Optimization Problem

Student **Lukas Jager,** Universität Bonn, Germany

Mentor **Radu Serban,** CASC

In this project, we considered the problem of determining a source in a time-dependent, advective-diffusive transport equation by given measurements of the concentration at certain locations. This problem is formulated as a constraint optimization problem where the objective function represents the difference between these measurements and the model prediction. The so-called reduced-space approach allows the reformulation as an unconstrained least squares minimization problem, which is typically solved by a Newton-type method. This entails the solution of linear systems involving the Hessian and the gradient of the objective function.

Adjoint sensitivity allows the computation of the gradient and the action of the Hessian on some vector with the same cost as the solution that the model system requires. But since the Hessian is not assembled, an iterative method for the solution of the linear system has to be applied. Due to the large condition number of the matrix, classical iterative methods provide only slow convergence. The use of a preconditioner can improve the convergence by changing the properties of the system matrix.

We implemented a multilevel preconditioner that consists of one or more coarser discretizations of the source. To solve the preconditioning system, it applies a full multigrid V-cycle. On the coarsest level, the discretization allows the assembly of the Hessian and solves the system by inverting the matrix. Our multilevel preconditioner enhances the convergence of the iterative solver and hence reduces the total number of iterations during the minimization.

The project will be continued by improving the performance of the preconditioner and considering different types of source discretizations that allow, for example, the determination of a time-dependent source.

Integration of Multivectors to Hypre

Student
Abram Jujunashvili, University of Colorado at Denver

Mentor
Charles Tong, CASC

The main project focused on the integration of multivectors to Hypre. To accomplish this, we created new data structure—parallel multivectors and the functions that are necessary to use them in Hypre. First, we built the sequential realization of several functions:

hypre_SeqMulti_VectorCreate
hypre_SeqMulti_VectorDestroy
hypre_SeqMulti_VectorInitialize,
hypre_SeqMulti_VectorSetDataOwner
hypre_SeqMulti_VectorSetConstantValues,
hypre_SeqMulti_VectorSetRandomValues
hypre_SeqMulti_VectorCopy
hypre_SeqMulti_VectorScale
hypre_SeqMulti_VectorAxpy
hypre_SeqMulti_VectorInnerProdGram
hypre_SeqMulti_VectorMultiScale
hypre_SeqMulti_VectorMultiAxpy
hypre_SeqMulti_VectorInnerProdDiag.

Based on these functions, we then constructed the parallel functions for parallel multivectors:

hypre_MultiParVectorCreate
hypre_MultiParVectorDestroy
hypre_MultiParVectorInitialize
hypre_MultiParVectorSetDataOwner
hypre_MultiParVectorSetPartitioningOwner
hypre_MultiParVectorSetMask
hypre_MultiParVectorSetConstantValues
hypre_MultiParVectorSetRandomValues
hypre_MultiParVectorCopy
hypre_MultiParVectorScale
hypre_MultiParVectorMultiScale
hypre_MultiParVectorAxpy
hypre_MultiParVectorMultiAxpy
hypre_MultiParVectorInnerProdGram
hypre_MultiParVectorInnerProdDiag.

The main problem we confronted while creating these functions was the presence of a multivector mask. The vectors are divided by active and passive parts—deflated vectors are passive and others are active. The mask is an array of ones and zeros. The i-th component is equal to one if i-th vector is active; otherwise, it is equal to zero. Because of the presence of a mask, all multivector functions became much more complex.

We changed the communications to reduce the number of messages for computing the inner product. We also changed the communications that are necessary for matrix-vector multiplication. These changes reduced the total number of messages and consequently, the time necessary for computations.

The main users of multivectors will be block iterative methods. One such method is the Locally Optimal Block Preconditioned Conjugate Gradient (LOBPCG) method for solving large eigenvalue problems. Based on general Hypre interface interpreter, we created an interface interpreter that gives us the ability to use abstract pcg code for multivectors.

In collaboration with Jean-Luc Fattebert (CASC), we solved several eigenvalue problems with very large matrices that came from real problems of electronic structure calculation. Coincidentally, we obtained similar results using LOBPCG and preconditioned (with multigrid preconditioner) gradient descent.

As a product of our joint work with Panayot Vassilevski (CASC), a paper titled, "On the Convergence Rate of the Conjugate Gradient Method Exploiting Arbitrary Initial Search Vectors," (P. Vassilevski, A. Knyazev, A. Jujunashvili) has been submitted.

I also collaborated with John Pask (H Division) and his summer student, Kristopher Andersen, to make multigrid preconditioned conjugate gradient methods for a new fortran code intended to solve large eigenvalue problems in electronic structure calculations.

TOC

Automatic Detection of Hot Spots in C/C++ Programs

Student **Alin Jula,** Texas A&M University

Mentor **Dan Quinlan,** CASC

My work at LLNL contributed to the automatic detection of hot spots in C/C++ programs. It contains two separate tools: the *outliner* and the *generic instrumentation tool*.

The outlining mechanism was implemented in the ROSE source-to-source translator, a project in CASC division overseen by Dan Quinlan. The outliner extracts portions of code from a source program into a standalone module and calls to the outlined modules are inserted in the remaining program. The outliner has the inverse functionality of the inliner. If outlining and inlining are performed on a program, the resulting program is semantically equivalent with the initial program. The user inserts special pragma statements that mark the portions of the code that need to be outlined.

The generic instrumentation tool allows programs to be instrumented with *any* instrumentation tools (TAU, PAPI, etc.) using a simple instrumentation format. This format is independent of the tool used, and contains the rules about where and what will be instrumented in the program. The "where" question is answered by using an expression in which the operands are pivotal concepts in a C++ program (loops – for, while, do-while, files, functions, scopes and function calls). The operators are the logical operators (,), and, or, not and in. The evaluation of an expression results in a collection of points in the program for which the instrumentation will be performed. The "what" question is answered by reading the user instrumentation statements from a separate file. This tool is useful for large programs where hand instrumentation could be extremely difficult and error prone.

The generic instrumentation tool is used for empirically detecting the hot spots in a program, while the outliner is used to extract the identified hot spots for further optimization.

Automatic SIDL Generation with ROSE

Student

Kirk Kelsey, University of Rochester

Mentor

Tom Epperly, CASC

The goal of my summer work was to allow for easier use of the Babel language interoperability tool. Babel lets a programmer access a software library implemented in one supported language to be called from any other supported language. Babel supports C, C++, Fortran 77, Fortran 90, Python, and Java. Its language interoperability is based on specifications written in the Scientific Interface Definition Language (SIDL). Typically a programmer would have to write the SIDL code by hand after writing the software library. My goal was to automatically generate the SIDL from the source code of the software library.

The generation of SIDL is being performed using the ROSE preprocessor generator. ROSE is a set of libraries that can be used to parse C source code and generate new code based on the structure of the original program. Although ROSE is typically used to generate more optimized C code, in this case it is being used to generate SIDL.

Once the SIDL is generated by ROSE, Babel must be invoked to generate skeleton code in each supported language to wrap the original library. This wrapping code facilitates access from other languages. Currently, the skeletal wrappers must be completed by the writer. The next stage of this work is to use a ROSE preprocessor to complete the skeleton code automatically. Once this is done, language interoperability will be achieved without any writing needed beyond the original library.

Large-Scale Atom Data Visualization

Student
Jason Kimball, University of California, Irvine

Mentor
Mark Duchaineau, CASC

Scientific visualization of the large sets of atoms produced by numerical physics simulations at the Lab presents specific challenges due to the large number and geometric nature of these atoms. Rendering data sets on the order of 10–100 million atoms requires the use of parallel rendering clusters that may not be able to provide interactive rendering speeds (60 frames per second with minimal latency), especially when the user's computer has a slower connection to the rendering cluster. Also, since atoms should be rendered as unique spheres, traditional, performance-improving, level-of-detail (LOD) algorithms based on geometry simplification are useless because simplifying the geometry of a large group of atoms removes the granularity of each atom.

Building upon the ROAM architecture developed by Mark Duchaineau, I implemented a perspective-based LOD system that renders progressive-quality views of the data. Each rendering becomes a texture on the wall of a cube that surrounds the "camera" viewing the data. From within this cube, the user is free to look in any direction and see data, at least in a low-resolution format. Meanwhile, the system renders progressively higher resolution views for the interesting regions of the data. Being able to request higher-resolution renderings for only limited areas of interest greatly reduces the rendering work. Furthermore, since the user has a view of the data at all times, this technique provides the desired interactive speeds, and the progressive resolution updates provide detailed information about the data.

In addition to overcoming the aforementioned problems with atom data, this solution works for data with very complex geometry and therefore, does not work well with traditional LOD algorithms.

Future work includes developing algorithms to optimize the selection of "interesting" regions, as well as using several different viewpoints and image warping to provide movable cameras.

TOC

Algebraic Multigrid for Maxwell's Equations with Variable Coefficients

Student

Nicholas Kridler, University of Colorado at Boulder

Mentor

Barry Lee, CASC

There are many difficulties in creating an efficient multigrid method for the curl-curl formulation of Maxwell's Equations. The key difficulty is the large near-null space of the curl operator. While there are several successful algebraic and geometric multigrid methods for this problem, performance degrades when there is a large variation in the material coefficients. The reason for this lies in the interpolation operator. In the approach by Hiptmair [3], the interpolation was constructed under the constraint that a multilevel commutativity complex is formed. However, under these constraints, the interpolation operators cannot handle divergence-free error components.

Ideally, one would like a complex that includes the divergence operator, but such an operator is extremely difficult to construct. Lee and Jones [4] constructed a scheme that relaxed on the multilevel commutativity constraint, allowing more freedom in the construction of the interpolation operators to handle both curl-free and divergence-free error components. In their approach, the null space of the curl operator is explicitly introduced into the set of equations. For this formulation, interpolation must be constructed for the edge and nodal elements. The nodal interpolation is constructed by box-mg [2] and the edge interpolation is constructed by AMGe [1].

Constructing the edge interpolation requires defining the weights for the boundary edges on the fine level and then solving a local problem for the weights corresponding to the interior. The ratio of the length of the fine edge to the length of the coarse edge is used to define the boundary weights. Unfortunately, this does not take the variable coefficients into consideration and will fail if there is too much variation. My project involved investigating ways to create edge interpolation based on the variable coefficients when only given a user-defined fine-grid matrix.

In Maxwell's Equations, the coefficient of interest is the magnetic permeability. Magnetic permeability is related to magnetic flux through a constitutive relation. Our goal was to create interpolation based on enforcing tangential continuity of the magnetic flux through the jump discontinuity of the coefficient. This requires assuming that all of the boundary weights are unknowns, and it is not clear that the problem will remain local. It is quite possible the fine edge I am trying to construct depends heavily on a neighboring element. The other problem is that the magnetic permeability is unknown, so it may not even be possible to enforce the constraint.

Another approach involves ignoring the fact that the coefficient is the magnetic permeability and looking at it as a stretching or shrinking factor. The effect of the coefficient can be seen as the transformation of the grid and weights could again be chosen by ratios of lengths. It is not clear that this transformation gives us any information we can use or if there are too many unknowns for it to be effective. Both approaches are still under consideration. Ideally, we would like a method that uses only values along a single coarse edge to create the corresponding fine grid elements.

References

[1] M. Brezina, A. J. Cleary, R. D. Falgout, V. E. Henson, J. E. Jones, T. A. Manteuffel, S. F. McCormick, J. W. Ruge, Algebraic multigrid based on element interpolation (AMGe), SIAM J. Sci. Comp., 22 (2000), 1570-1592.

[2] J. E. Dendy, Jr., Black box multigrid, J. Comput. Phys., 48 (1982), 366-386.

[3] R. Hiptmair, Multigrid methods for Maxwell's Equations, SIAM J. Numer. Anal., 36 (1999), 204-225

[4] J. E. Jones and B. Lee, A Multigrid Method for Variable Coefficient Maxwell's Equations, SIAM J. Sci. Comp., submitted.

TOC

IMAGE Pipeline

Student **Brian Lum,** University of California, Berkeley

Mentor **Amber Marsh,** EEBI Division

I worked on a group project called the Integrated Molecular Analysis of Genomes and their Expression (IMAGE) Pipeline Project. I helped the group meet deadlines by taking on a variety of tasks that otherwise would have had to become post-production projects or would have forced the deadline to be pushed back.

My first task was setting up barcode printers. The printers required a great deal of complicated configuration, and it was my job to work with the printer company and with the Laboratory's information technology (IT) group to ensure that everything was set up to meet our needs. I also determined how to program the printers and how to design a printer command generation module that integrated with the rest of the Laboratory's Web-based interfaces.

My second task was designing and implementing a Web-based interface to maintain the personnel database related to the project. I also retrieved and formatted the data that I eventually used to populate the initial personnel database.

In a similar side project for IMAGE, I worked with the project's Principal Investigator (PI), Christa Prange, to retrieve outdated information from an old database, update it, reformat it, and enter it into the new system. I then designed and implemented a Perl script that is run nightly via CRON to display the information.

Numerical Solution Methods and Error Analysis for the Neutron Transport Equation in Absorbing, Monoenergetic, Non-Scattering Media

Student **Eric Machorro,** University of Washington

Mentor **Britton Chang,** CASC

We considered the local truncation errors and estimated global errors for the exponential-characteristic (EC) method for solving the 2D monodirectional, non-scattering, Boltzmann neutron-transport equation. We showed that under ideal circumstances, the local error is second order, yet globally, the method is only first-order accurate. Test problems were evaluated, and in some respects, the EC method performed better on problems with discontinuous material interfaces. An analytic counter-example was explored demonstrating at best O(h) accuracy.

As an alternative, we developed the Boundary Exponential Log Characteristic (BELCh) method. The method is positive and locally satisfies the conservative form of the equation. Estimates for local truncation errors and global errors for the method are under development.

For strictly positive solutions that have L-continuous derivatives, we show that BELCh is locally Lth order accurate, yet globally, the method is only L–1-order accurate. Examples and test problems were explored, and in some important respects, BELCh outperforms the EC method.

We also proved that the Petrov-Galerkin discretization of the monodirectional neutron transport equation is convergent. This was demonstrated by proving that the inverse of the matrix for the system is bounded by a number that is independent of the order of the matrix. Therefore, the global error of the Petrov-Galerkin approximations of the transport equation has the same order as the local truncation error.

Out-of-Core Visualization of Climate Modeling Data

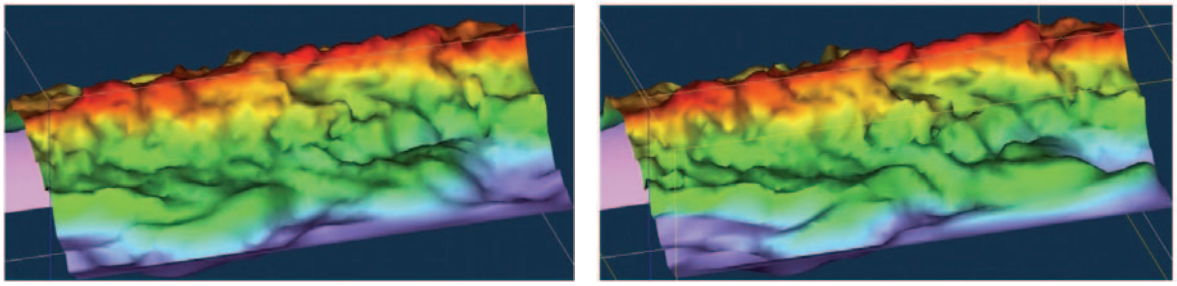
Student **Anna Majkowska,** University of California, Riverside

Mentor **Valerio Pascucci,** CASC

My research activity during last summer has been focused on enhancing the ViSUS IDX streaming infrastructure. The work has been centered around two main tasks:

- Extension of the data extractor in the streaming infrastructure to allow synchronized retrieval of multiple scalar fields.
- Development of a new out-of-core isosurface extraction algorithm that takes advantage of hierarchical Z-order data layout. The scheme allows for view dependent refinement and avoids storing explicit data structure for the triangles in the isosurface. In particular, the vertices of the isosurface are computed by a vertex program that takes a tetrahedron T and then renders the portion of isosurface contained in T.

The main test bed for the new features has been the 3D image data obtained from the IMPACT climate modeling simulation. With the current enhancement, one can connect the ViSUS streaming framework directly to the simulation, and while the simulation is running, visualize isosurfaces of the ozone concentration pseudocolored by temperature distribution (see figure below). The synchronized streaming allows for the introduction of streaming computation of derived scalar fields, which is a crucial component of any complex visualization environment for large scientific data.



Two time steps of the IMPACT climate modeling simulation visualized with an isosurface of the ozone concentration pseudocolored with second scalar field.

TOC

Code-Independent Analysis Tools for Physics Simulation Codes

Student **Robert Dean Malmgren,** Northwestern University

Mentor **Scott Brandon,** AX Division

It is becoming increasingly important for computational physicists to be able to analyze the accuracy of their codes during calculation as well as in postprocessing. While visualization techniques are readily available, simulation precision cannot be judged visually. Furthermore, it is impractical to visualize the problem during computation to judge accuracy. Code-independent analysis methods, which offer both in-line and postprocessing capability, are therefore developed to give computational physicists a means of rigorously analyzing any code's performance.

The developed tool quantitatively measures directional symmetry, relative error contours, and relative shock front location based on the computed and analytic solutions. Inline implementation of this tool gives computational physicists the ability to have their codes adapt and self assess throughout, for example, a parameter optimization experiment. In postprocessing mode, these techniques allow computational physicists to examine error patterns in their codes and ultimately facilitate code development and fine tuning.

The following figures illustrate the tool's capability compared to KULL's performance against the predicted analytic solution of the Noh Problem. Figure 1 demonstrates the ability of the symmetry analysis tool. Figure 2 shows the shock front tool's use. Both tools can be used to quantitatively describe asymmetry when the asymmetric profiles are chosen. In this case, the profiles along the axes are asymmetric.

Future work on this project could extend the tool's capability to include an asymmetry finder. Additionally, this tool could be altered to quantitatively compare code performance against experimental data or another code's performance.

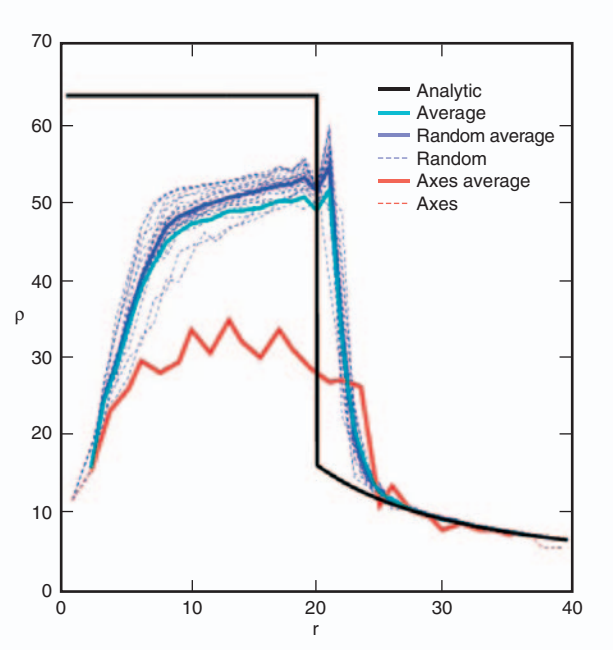


Figure 1. Symmetry analysis of KULL solution to Noh problem

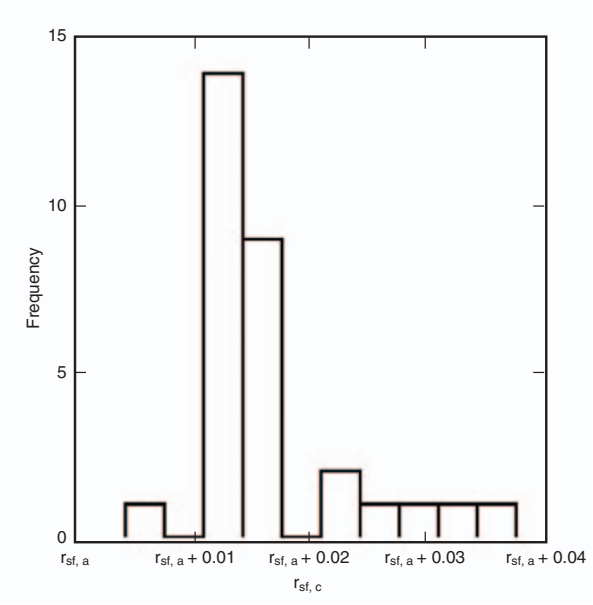


Figure 2. Histogram of relative shock front position of calculated profiles

Development of a Graphical User Interface for the Virtual Beam Line (VBL) Simulation Software

Student

Peter Manning, University of Arizona

Mentor

Glenn Goderre, NIFE Division

The purpose of the project was to develop a graphical user interface (GUI) to interact with and utilize the already-developed VBL software. Major accomplishments include: a method for viewing and editing part properties (parts of the NIF laser); the ability to open any tabulated aberration file, gain file, or beam dump file and view its distribution in 2D and 3D; a mechanism for inserting beam dumps anywhere along the path of the beam and also the ability to remove them; the ability to search for a part and have its information displayed; the ability to recall previous runs and their summaries. The GUI displays the output summary from a VBL run in a readable and configurable format.

I learned more Java techniques and design patterns than I thought possible in a single summer. After 10,000 lines of code and 160 classes, there exists a program that enables users to easily work with the VBL simulation software. Although the GUI will change in many ways in the next few months, these accomplishments provide the foundation for future changes and advancements.

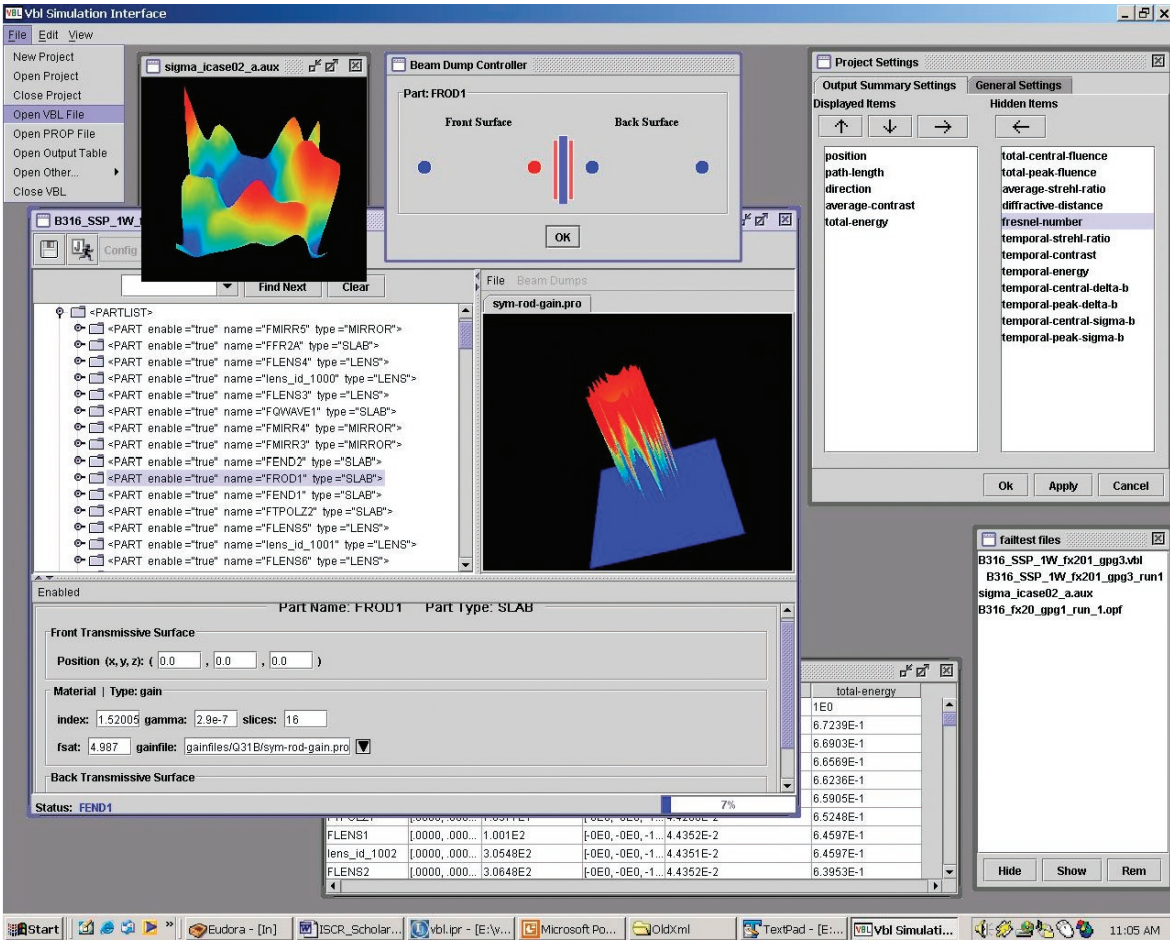


Figure 1. Screen shot of VBL GUI and most of its capabilities during a simulation run

Time-Varying Reeb Graphs for Space-Time Data

Student
Ajith Mascarenhas, University of North Carolina at Chapel Hill

Mentor
Valerio Pascucci, CASC

Physical processes that are measured over time, or modeled and simulated on a computer, can produce large amounts of data that must be interpreted with the assistance of computational tools. Such data arises in a wide variety of studies, including computational fluid dynamics, oceanography, and climate modeling. The data typically consists of finitely many points in space-time and a measured value for each.

Graphical visualization, often through level sets or isosurfaces of a continuous function, is useful for interpreting the data. A level set consists of all points in the domain whose function values are equal to a chosen real number 's'. In three dimensions, this is generically a surface that is then displayed. By varying 's', we can study the variation in the data. Topological features of the level sets, such as connected components, handles, and voids, can be important aids in interpreting the data. By encoding the evolution of these features, the Reeb graph compactly represents topological information of all level sets. As we pass through time, the Reeb graph goes through an evolution of its own, undergoing structural changes at birth-death points and at interchanges of critical points. The evolution of the Reeb graph thus represents a 2-parameter family of level sets. We suggest that this 2-parameter family, encoded in a compact data structure, is a useful representation of space-time data.

I developed software that computes time-varying Reeb graphs specialized for the case of 2D scalar fields. I first worked on the mathematical foundations of Morse functions and then addressed the practical issues that allow a robust implementation. A central challenge in this work is defining a set of consistent rules to handle the degeneracies present in real-world data. These policies bridge the gap between the theory for smooth functions and the practical world of sampled fields obtained from simulations or experiments. The software development is in progress and will be finalized in the next two months at the University of North Carolina at Chapel Hill.

I also started the next stage of this research, involving the construction of multiresolution Jacobi sets, which are the 1-manifolds that describe the trajectories of critical points over time. I implemented the initial ideas in a software prototype that visualize the evolution of a 2D scalar field. In this tool, the Jacobi sets are first presented at full resolution and are then incrementally simplified to highlight the most important features in the data. The preliminary tests with combustion data sets show promising results. The figure below shows the Jacobi set of the combustion data with three levels of simplification, starting from full resolution at left.

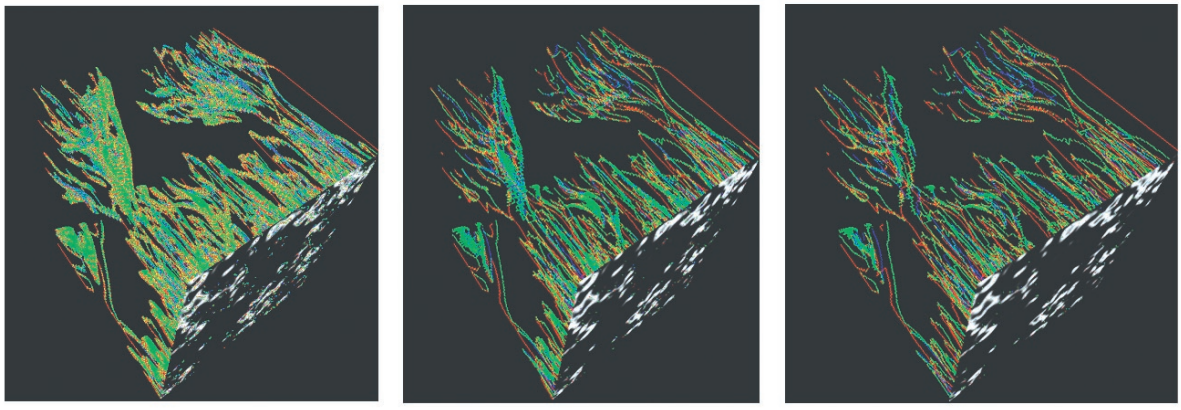


Figure 1. Three levels of simplification starting with full resolution (left) depict a Jacobi set.

Ultra-Wideband (UWB) Token Ring Simulation & Security Needs of Sensor Networks at LLNL

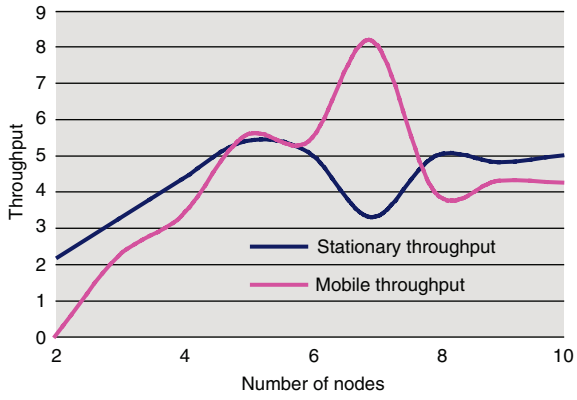
Student **Tammara Massey,** Georgia Tech / University of California, Los Angeles

Mentor **Terry Brugger,** NAIC

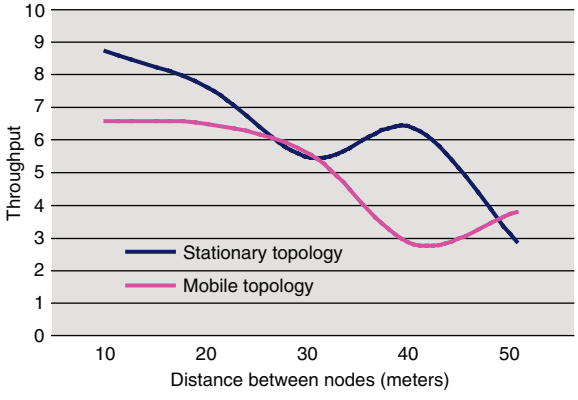
Ultra-wideband is an emerging wireless technology that permits covert and reliable transfer of data. It is essential to investigating networking protocols, such as token ring, that allow for quality of service and bounded latency for ultra-wideband radios. To explore the limitations of the ultra-wideband token ring protocol, throughput was measured with varying levels of mobility and with various quantities of nodes. With a larger number of nodes, the throughput increased because each node could transmit data to more possible nodes. The throughput for nodes with mobile topologies was lower than for nodes with stationary topologies due to the additional overhead of mobile nodes joining and leaving rings. The simulation also showed how a larger distance between the nodes and a larger topological area decreased the throughput. The simulation proved that the token ring protocol provided the desired throughput and quality of service for ultra-wideband radios transmitting multimedia. I created a poster and a technical paper describing my research and suggesting further work. With the aid of my mentor, this paper will also be submitted to a conference.

Security in Sensor Networks

Several wireless sensor technologies exist at LLNL that gather data for environmental or defense applications. Unfortunately, wireless technologies are more vulnerable to security attacks because the data is broadcast. The ubiquitous nature of broadcasting allows adversaries to eavesdrop or inject messages into the wireless network more easily than in a traditional wired network. After describing the current sensor networks at LLNL, I analyzed the security vulnerabilities of sensor networks. I presented security solutions that need to be implemented in current and upcoming projects. I also proposed future areas of research that need to be explored to ensure that LLNL sensor applications remain safe and reliable. I wrote a technical paper describing this research.



Throughput for Ultra-Wideband Token Ring Simulation with an Increasing Number of Nodes



Throughput for Ultra-Wideband Token Ring Simulation with an Increasing Distance Between Nodes

PerfTrack

Student **Kathryn Mohror,** Portland State University

Mentor **John May,** CASC

PerfTrack is a tool to help scientific programmers answer difficult questions about application performance, given that the source code, build parameters, runtime environment, and hardware vary over time. PerfTrack was developed to explore technologies in parallel performance measurement, modeling, analysis, and prediction. Performance data and the associated environment data is being stored in a relational database. This database provides a foundation to build analysis tools that are scalable to large numbers of threads and are capable of comparing multiple executions. Our goal is an automated tool that will gather, store, and analyze data, in order to encourage their use in the software development cycle.

This project was motivated by needs in the high-performance computing community. Code teams need to have an organized way to understand how changes in hardware, build and runtime environments, and source code affect application performance. Machine purchasing teams need a way to predict how future machines will perform so that they can select machines with characteristics that are best suited to the codes that are important to them. PerfTrack addresses these needs by researching solutions for problems that prohibited such tools in the past, such as how to compare possibly disparate performance data, and given a large amount of performance data, how to present only relevant information to the PerfTrack user. In future work, we plan to investigate other research problems, such as how to provide automatic correlation of performance trends with machine, software, and environmental characteristics.

Summarizing Network Traffic with Information Visualization

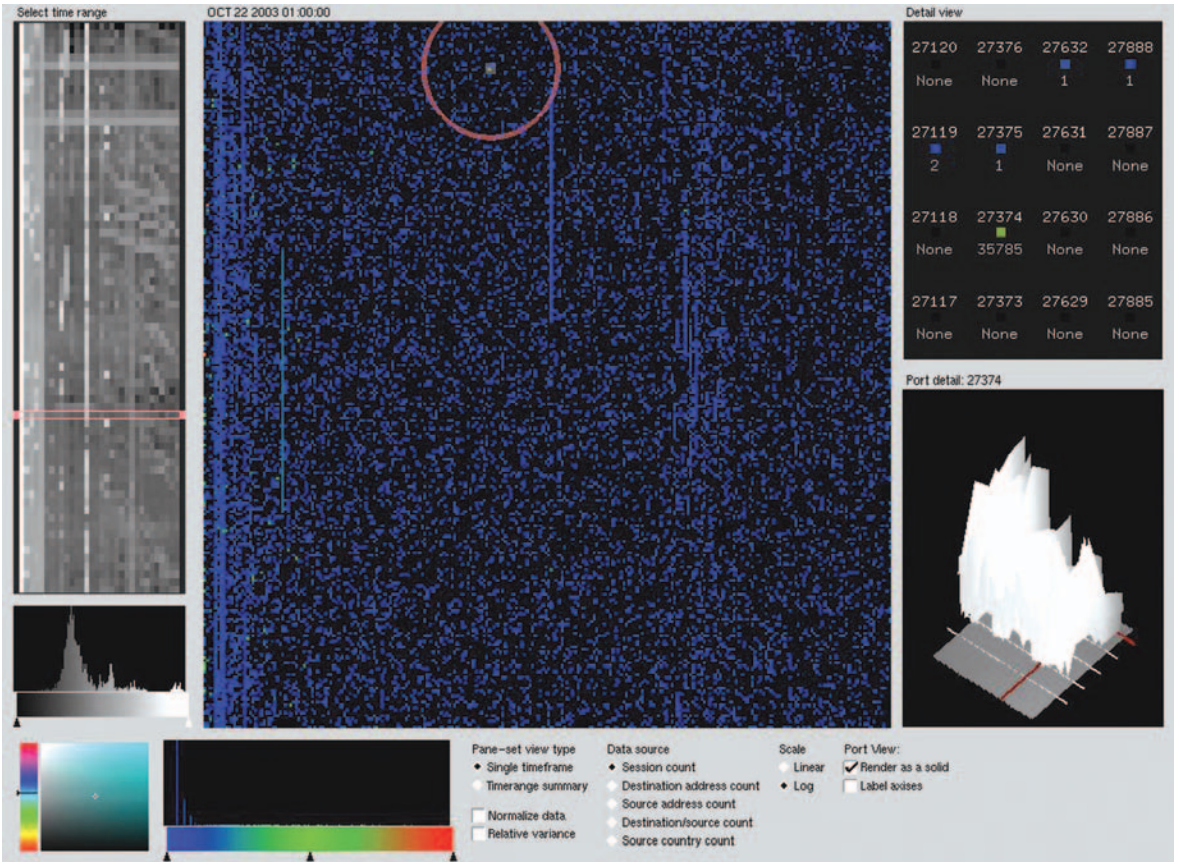
Student

Christopher Muelder, University of California, Davis

Mentor

Marvin Christensen, NAIC

Lawrence Livermore National Laboratory and several other Department of Energy sites have sensors on their networks that collect vast amounts of data. The quantity of data collected is too large to go through by hand, and automated analysis is not yet completely reliable. A program was developed to display a high-level visual summary so that suspicious behavioral patterns could be easily and rapidly detected. I worked on porting the program to Mac OS X, changing the layout to be more intuitive, improving existing capabilities, adding new capabilities, and fixing many small internal issues. The process of porting the program to OS X gave me the opportunity to get accustomed to the operating system and the existing code, and to learn how programs have to be changed to work on OS X. The layout of the program was rather unintuitive, flowing in essentially a counterclockwise direction. The new layout flows from left to right and is smaller overall so that it will work on screens as small as 1024x768. One of the visualizations was labeled in order to make it easier to use, and the histograms were improved to increase their color association and make it easier to edit custom gradients. New capabilities include the ability to normalize the data so that data from different parts of the day can be compared, and the ability to calculate variance relative to the values averaged across time periods, as well as an absolute variance.



PortVis is a graphical interface that allows computer security analysts to quickly sort thought large amounts of connection data to identify interesting abnormalities that warrant further investigation. Using this tool is a quick way to filter out normal network activity and focus on suspicious activity regardless of the variance of malicious code.

TOC

Software Failure Risk Analysis

Student **Spencer Nielsen,** Brigham Young University

Mentor **Darrel Whitney,** CADSE

My assignment was to create a Web-based application to enable project managers at the Lab to grade and assess their projects for likelihood of software failure and the risk consequences associated with failure. After the likelihood of failure and risk consequences of a project have been reported, the project is assigned a calculated grade. This grade is then used to consider practices and policies to implement in the project and to make other management decisions. I also wrote an administrative interface to allow directorate-level management to customize the evaluation process for projects in their directorates. I used the PHP/MySQL development platform to accomplish this task.

A Space-Time Sparse Grid Approximation Space

Student **Daniel Oeltz,** Rheinische Friedrich-Wilhems Universität Bonn

Mentor **Panayot Vassilevski,** CASC

In this project, we developed and implemented new approximation spaces for parabolic problems, so-called space-time sparse grid spaces. With classical space-time discretization spaces, using a uniform grid in d-dimensional space with N-grid points in each coordinate direction usually needs $O(N^{-(d+1)})$ or $O(N^{-(d+2)})$ degrees of freedom. Those spaces will reduce the number of degrees of freedom to $O(N^{-d})$ or $O(N^{-(d+1)})$. In the first step, we analyzed the approximation rates of those spaces. We were able to prove for special function classes that, besides the tremendous reduction of the degrees of freedom, the space-time sparse grid spaces provide nearly the same approximation rate as classical space-time Finite Element spaces.

In opposition to classical sparse grid approaches, we are using a tensor product construction of a d-dimensional multilevel basis in space and a one-dimensional multilevel basis in time, which makes it possible to handle even complicated domains in a very easy manner. Here, we construct the spatial multilevel basis from a spatial Finite Element discretization via AMG coarsening and a least square approach. We implemented this approach, including adaptivity, and performed several numerical experiments showing that this approach is able to approximate space-time functions with great accuracy while reducing the number of the overall degrees of freedom significantly compared to Finite Element spaces. It turned out that by applying adaptivity, this approach provides good results, even for highly singular functions.

Reviewing Vista

Student

Christopher Olson, University of California, Santa Cruz

Mentor

Robert Cooper, DCOM

Computation loops in ALE3D involve a fixed notion of an index map. I implemented two easy methods to abstract this indirection in an effort to reduce table lookups and better respect cache lines. Unfortunately, the cost of both techniques, especially from register spilling due to additional register pressure, outweighed the anticipated benefits. Other techniques may still allow us to clarify code and reduce the runtime of computation loops, but they will require more effort to build, and they may involve significant changes to the existing source code. I also rewrote ALE3D's Vista library to get a cleaner and faster implementation.

Subspace Detectors

Student

Timothy Paik, University of California, Berkeley

Mentors

Shawn Larsen & Dave Harris, ISCR

When handling seismic signals, it is often necessary to correlate different signals against each other. One of the tools used to correlate seismic signals is the subspace detector. Subspace detectors calculate template signals from events and correlate them against other events. Subspace detectors are useful because they return correlation values independent of the amplitude of the signal. This behavior is helpful for two reasons. First, they do not return misleading large correlation values when receiving large signals. Second, they can find correlating seismic events that might be buried under the noise of other unimportant signals.

My assignment this summer was to write Java classes designed to improve the signal-processing engine of the subspace detector. I wrote several classes to represent complex signals and several other classes to process complex signals. My main goal was to improve the overall speed of the subspace detector. I achieved this goal through a change in the algorithm and the development of a quick Fourier transform. The subspace detector now runs about 3.5 times faster than before, allowing for greater amounts of subspace detectors to be deployed in the field.

TOC

Streaming Pointsets

Student

Sung Park, University of California, Davis

Mentor

Peter Lindstrom, CASC-ISCR

The increasing amount of point data generated by different applications is posing a major challenge in data processing. Typically, such large data sets not only exhaust the main memory resources of common desktop PCs but can even exceed the 4 GB address-space limit of a 32-bit machine. To process data sets that do not fit in main memory, out-of-core (external-memory) algorithms are used.

Dealing with large data sets for the purpose of scientific visualization is an ongoing area of research. For large point data, there has not been much work to address this problem. One ideal approach for dealing with large point data is a streaming model. In a streaming model, data is streamed into the main memory and retired as soon as it is finished processing. My supervisor's previous work has shown that this approach works well for large meshes.

My work this summer has been trying to come up with an optimal approach to do stream processing on point-based data sets. The problem proved to be a lot harder than anticipated as many different approaches we tried gave poor results. The problem is difficult because pointsets do not provide connectivity information, thus finding even neighboring information requires certain assumptions about the data that has been streamed into the machine. The approach I recently investigated has shown some good results and deals with finalizing partitions of space by using a binary partition tree. The work is still in its beginning stages but preliminary results have been good, and we are hoping to further investigate this idea.

Internet Ballistics: Retrieving Forensic Data From Network Scans

Student **Bryan Parno,** Harvard University

Mentor **Tony Bartoletti,** Computer Incident Advisory Capability (CIAC)

The typical network receives millions of hostile probes every day, and a significant portion of these probes are network scans. During a network scan, the attacker sends connection requests to every possible network address and listens for replies indicating the presence of a (possibly vulnerable) computer. Since a network scan often serves as a precursor to an attack, reliable identification of scanners can significantly enhance cyber-security. Furthermore, the ability to map adversary hierarchies and correlate the attacks with events in the real world contributes to counterintelligence work.

For a variety of reasons, source IP addresses fail to provide the necessary identification information. However, analyzing packet arrival timing data reveals highly distinctive patterns that may correlate with the attacker's choice of tools, physical platform and/or network location (Figure 1). By selecting data transforms conducive to periodic analysis, we can use wavelet techniques to achieve more than 1,000x compression ratio while still preserving the essential features (Figure 2). Initial experiments indicate our methods consistently identify patterns in the data. In future work, we plan to perform controlled scans using common network scanning tools from multiple locations to refine our identification techniques, allowing us to reliably fingerprint network scanners, without relying on the source IP address.

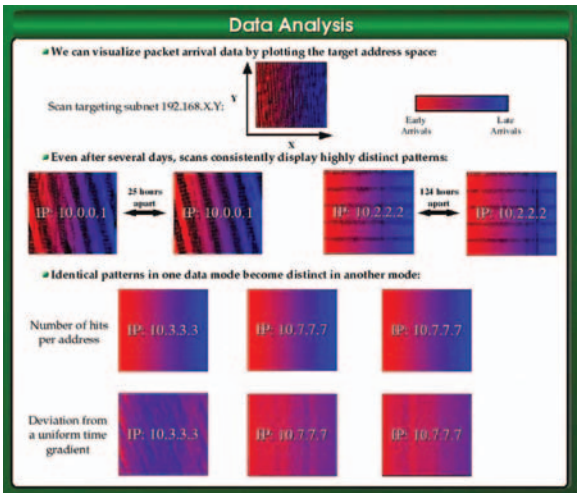


Figure 1. Class-B subnet plots of various hostile packet statistics (packet arrival times, hits per target address, and others) reveal multiscale structure often characteristic of source."

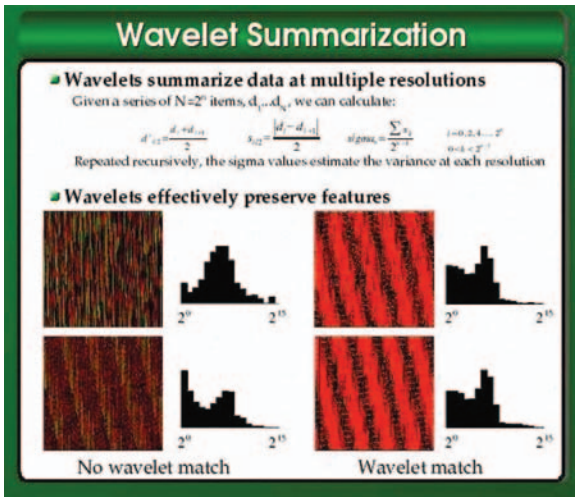


Figure 2. Wavelet scalograms reliably capture and record multiscale structure in scan-traffic statistics, providing compact "scan-behavior signatures" for identification purposes.

SLURM for BlueGene/L

Student **Dan Phung,** Columbia University

Mentor **Moe Jette,** CASC

My duties this summer were related to the incoming system installation of IBM's BlueGene/L (BGL), currently the world's fastest computer. I was given the task of porting the code that currently manages the resources of the other large cluster systems for the BGL system. My other responsibility was to design and implement a partitioning and wiring algorithm that divides the system into portions to allow the system to be shared among users and efficiently utilizes the system.

Software Vulnerability Taxonomy Consolidation

Student **Sriram Polepeddi,** Carnegie Mellon University

Mentor **Noel Tijerino,** IOAC

The goal of the software vulnerability taxonomy consolidation project is to address the need for a universally accepted vulnerability taxonomy that classifies vulnerabilities in an unambiguous manner. In today's environment, computers and networks are increasingly exposed to a number of software vulnerabilities. Information about these vulnerabilities is collected and disseminated via various large publicly available databases, such as BugTraq, OSVDB, ICAT. These databases do not cover all aspects of a particular vulnerability and lack a standard format, making it difficult for end-users and administrators to easily combine and compare various vulnerabilities. A central database of vulnerabilities has not been available until today for a number of reasons, including non-uniform methods by which current vulnerability database providers receive information, disagreement over which features of a particular vulnerability are important and how best to present them, and the nonutility of the information presented in many databases.

A consolidated vulnerability database (CVDB) was implemented that clearly communicates information between the vulnerability discovery agent and end users of that data. Data from BugTraq, OSVDB, ICAT, Secunia, Computer Associates' Vulnerability Information Center, and Mitre's CVE was fed into a database. These sources required extracting information directly from Web pages, XML, CSV and Tab-delimited files. Once loaded into a database, the data most unambiguously representing an aspect of a vulnerability was loaded into the main CVDB tables and linked using CVE IDs. While representing a smaller subset of the total available data, the information now possesses a precise and repeatable structure.

TOC

Performance-Oriented, Privacy-Preserving Data Integration

Student
Raymond Pon, University of California, Los Angeles

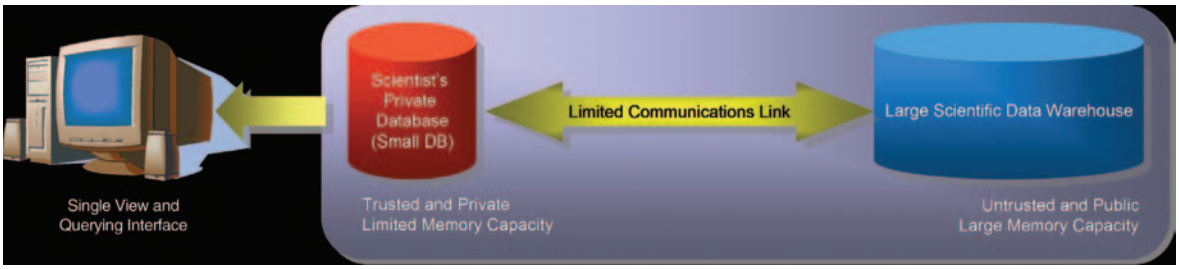
Mentor
Terence Critchlow, CASC

Data is often generated or collected by various parties, and the need to integrate the resulting disparate data sources has been addressed by the research community. Although the heterogeneity of the schemas has been addressed, most data integration approaches have not yet efficiently addressed the privacy requirements imposed by data sources. Without strong privacy guarantees, scientists often refuse to share data with other scientists for reasons such as subject/patient confidentiality, proprietary/sensitive data restrictions, competition, and potential conflict and disagreement.

When sharing scientific data, privacy quickly becomes an issue. Suppose that a scientist wishes to perform a query across a table in his private database and a table in a public data warehouse in the most efficient manner possible. Ignoring privacy restrictions, the problem is reduced to a distributed database problem that can be solved by shipping the scientist's table to the data warehouse and performing the join at the warehouse. However, if the scientist's data set is proprietary, it cannot be sent verbatim to the data warehouse. The naive solution is for the scientist to download the entire data warehouse to his local machine and perform the query there. But to do so would be prohibitively expensive if the data warehouse is very large or the communications link is limited.

To address this problem, we augment the well-known semi-join framework, "hiding" the actual values of the join column of table *R* by hashing them and including additional artificial values. Our hash/noise method approach uses a set of fixed hashing and fake hash values to control the amount of uncertainty in the identity of the join column values in table *R*, thereby controlling the level of privacy loss incurred. By sacrificing a small fraction of privacy, this method incurs significantly less transmission costs than the naive solution. Relative information gain is used as the measure of privacy loss.

Experiments were done to evaluate the effectiveness of this approach. The results of the experiments will be published at the next appropriate conference.



Target architecture for privacy preserving queries

Out-of-Core Parameterization and Remeshing of Surfaces

Student **Serban Porumbescu,** University of California, Davis

Mentor **Mark Duchaineau,** CASC

The goal of this project is to perform remeshing and parameterization of surfaces too large to fit into the main memory of conventional computers.

Our approach creates a hierarchy of fine-to-coarse resolution surfaces via repeated removal of an independent set of vertices. The independent set consists of non-neighboring vertices. This construction process results in a hierarchy of topologically equivalent surfaces and a robust one-to-one and onto mapping between these surfaces. Consequently, a point on a surface within the hierarchy can be mapped uniquely to any other surface in the hierarchy.

With this hierarchy at our disposal, we perform remeshing by regularly subdividing the coarsest level of the hierarchy and mapping the newly inserted vertices to a higher-resolution surface within the hierarchy. This mapping is generally not smooth and the introduced elements are highly distorted. To ameliorate this problem, we apply a smoothing filter to our remesh surface that regularizes elements with respect to a particular level of the hierarchy.

We are currently exploring several approaches to smoothing of the remesh surface with respect to a surface in the hierarchy. The first approach involves smoothing the remesh surface with respect to a portion of the target surface embedded in the plane. An alternative approach involves smoothing with respect to a locally constructed scalar field.

Future work involves mapping our algorithms and data structures to their out-of-core counterparts. While not trivial, our highly localized algorithms and data structures should help facilitate a smooth transition to an out-of-core approach.

TOC

Intrinsic Function Testing and an OpenMP Runtime Library

Student

Ashley President, Carnegie Mellon University

Mentor

Mike Kumbera, DCOM

The purpose of the project is to add necessary and useful features and to test existing and new features of GNU gfortran, an open source Fortran 95 project. The feature extensions that needed to be added are OpenMP, Application Program Interface (API), and Cray Pointers. Along with the added features, the current features of Fortran 95 need to be thoroughly tested.

Some of the intrinsic functions, required by Fortran 95 specifications, needed to be tested in gfortran. Once the tests were created, they were used to check the functionality of several preexisting Fortran 90/95 compilers. Forty-eight different functions were tested.

OpenMP is an Application Program Interface (API) that can direct multi-threaded, shared memory parallelism in Fortran, C and C++. OpenMP calls work with a set of simple directives that explicitly control the programs parallel activities. Along with directives, the OpenMP standard also specifies a set of runtime library calls that can be used to determine the system state. The OpenMP Runtime Library is designed to work with both C/C++ programs as well as Fortran Programs. The backbone of the runtime library is written in C and uses POSIX (pthreads)to implement threads.

While many of the gfortran features were tested, many other features within the compiler still need testing. The OpenMP runtime library stills needs to be tested on a larger scale. The C/C++ compilers need to be updated to recognize OpenMP calls.

TOC

AS2TS Web site

Student **Davinder Rama,** California State University, Sacramento.

Mentor **Adam Zemla,** EEBI

The Amino-acid Sequence into Tertiary Structure (AS2TS) and Local Global Alignment (LGA) systems provide a capability to create 3D models of analyzed proteins, detected signature regions and other biologically important fragments of protein structures. The goal of my project was to develop a publicly accessible Web site that will allow submission of queries to the selected local services:

Protein Structure Analysis services (LGA):

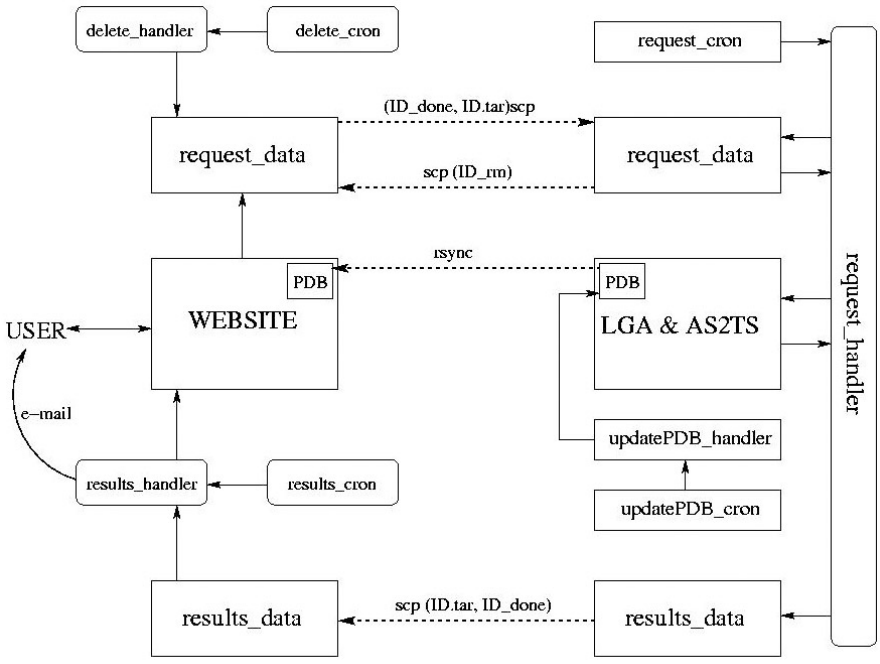
- Protein structure comparison facility — Finds 3D similarities in protein structures. Allows the submission of two 3D protein structures for structure comparative analysis.
- LGA–AS2TS Models structure comparison — Allows the submission of the protein structure in Protein Data Bank (PDB) format for LGA comparison with models generated by AS2TS system.
- LGA–PDB chains structure comparison — Allows the submission of the protein structure in PDB format for LGA comparison with chains from the PDB list.

Protein Structure Modeling services (AS2TS):

- Sequence-structure Alignment into Tertiary Structure — Generates a tertiary structure (3D model) for a given sequence-structure alignment model.
- AS2TS Service — designed to facilitate the modeling of tertiary structure (3D model) for a given sequence of amino acids.
- AS2TS Model Builder — An expanded version of AS2TS that allows submission of a set of protein sequences for high throughput protein structure modeling.
- Local Sequence Library for AS2TS system — Create and deposit Libraries in FASTA formatted protein sequences or combine existing libraries.
- Local Structure Library for AS2TS system — Deposit PDB structures in PDB format into the local PDB.
- SCWRL (Side Chain placement With a Rotamer Library) — side chain builder for AS2TS system. For a given protein structure SCWRL calculates new conformation of side chain atoms.

This project required establishing efficient data exchange (query<->results) between the web-server and the computation servers. An overview of the system follows:

The AS2TS website: http://as2ts.llnl.gov



The figure diagrams the interfaces and the data flow between the major modules of the developed system for data exchange between the "WEBSITE" server and the computation "LGA & AS2TS" server. To process user queries, three types of modules were developed—request_handler, results_handler, and delete_handler. Each module is invoked by a cron, and the data exchange is done using scp (secure copy, a remote file copy program).

TOC

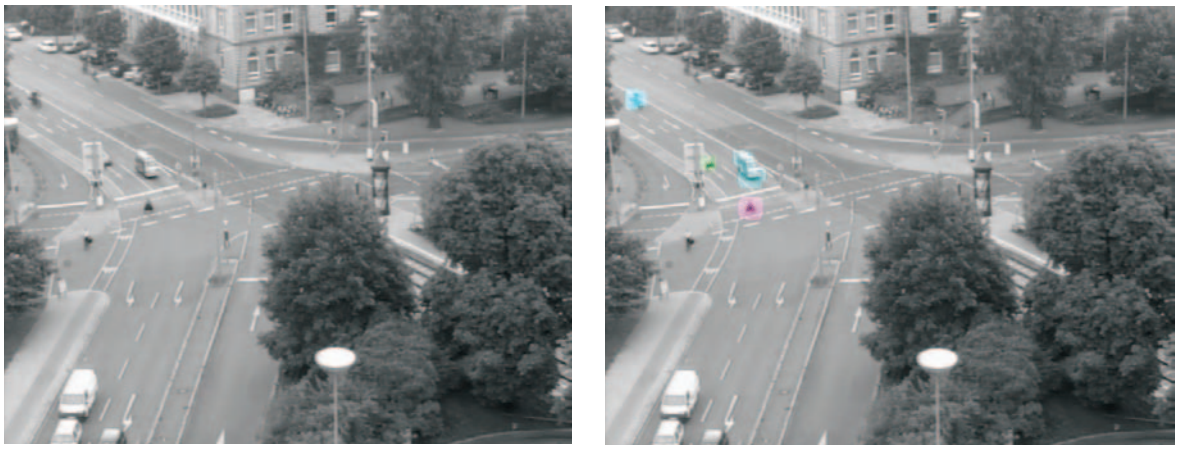
Variance-Based Feature Tracking

Student **George Roberts,** Georgia Institute of Technology

Mentor **Chandrika Kamath,** CASC

Many modern computer vision applications require the tracking of objects of interest. These objects of interest are sometimes stationary, which can complicate attempts to identify them. The goal of this project is to intelligently identify vehicles in video sequences under less than optimal conditions. While it is assumed that the video sequences are obtained from a stationary camera, algorithms robust to small camera movements are desirable. The algorithm used should also have low computational complexity. This paper focuses on both the recognition of objects of interest, as well as object correlation from one frame to the next.

Many methods, such as Background Subtraction and Corner Detection, have been proposed to perform this task. Background Subtraction identifies moving objects by comparing each video frame with a calculated background model. Alternately, Corner Detection finds the right angles in the images and tracks their movement to represent the moving objects. In this project, we propose an alternate variance-based algorithm. The speeds, locations, and intensity distributions of the objects are then used to assist in frame-to-frame correlation. The effectiveness of this variance-based algorithm is also compared to that of Background Subtraction and Corner Detection.



A scene from a video of a traffic sequence (left image), with the moving objects highlighted (right image). The moving objects were obtained using the variance-based method.

Multigrid Prolongation Based on Sharp Convergence Theory

Student **Rolf Ryham,** Pennsylvania State University

Mentor **Rob Falgout,** CASC

I was assigned to a project initiated by Visiting Scholar Prof. Ludmil Zikatanov and CASC faculty Panayot Vassilevski under the supervision of Rob Falgout. The project entailed the study, implementation, and presentation of an optimal prolongation algorithm for a two-level method based on a theory developed by the above three authors. An expose of the algorithm will be given below. Prolongation is one of three major user- or problem-defined components (prolongation, relaxation, coarsening) in a multigrid algorithm and is thus an integral part to the Linear Solvers Group's development of scalable solvers for large linear systems.

The nonconvexity of the algorithm makes it unclear whether certain key approximation steps are convergent or even stable, and we did not succeed in implementing a convergent algorithm for one- or two-dimensional problems. However, many modifications and test cases have been postulated for further development and will be implemented in the future. The prolongation operator P is mapping from the lower-dimensional coarse space to the high-dimensional fine space whose span determines the coarse basis. The coarse basis should reasonably represent the error after a smoothing operation, the essence of which may be expressed in the sharp bound:

$$\rho(E) = 1 - \frac{1}{K}, \quad K = \sup_{v \in R^n} \frac{\left\| (I - \Pi_{\tilde{M}})v \right\|_{\tilde{M}}}{\left\| (I - \Pi_A)v \right\|_A}$$

Choosing an optimal prolongation means to minimize K with respect to those P that satisfy a (linear) sparsity and an approximation constraint. The existence of a P is guaranteed but may not be unique. The algorithm postulates that an optimal P is a fixed point of the following iteration

1. Choose a high frequency v to represent the coarse grid complement.
2. Choose P that minimizes $\left\| (I - \Pi_{\tilde{M}})v \right\|_{\tilde{M}}$.

These steps are more difficult than the original matrix inversion. However, performing these approximately is feasible at the cost of knowledge of stability.

The AST Query Mechanism and the C/C++ Graphing Mechanism

Student

Andreas Sæbjørnsen, University of Oslo

Mentor

Dan Quinlan, CASC

I worked on two sub-projects of the ROSE project in cooperation with my supervisor Dan Quinlan. The sub-projects were the AST Query Mechanism and the C/C++ Graphing Mechanism, which are tools to extend and add capabilities to ROSE. They aim at increasing the productivity of the library developer using it.

The AST Query Mechanism is introduced as a building block when implementing algorithms and conditions in the ROSE abstract syntax tree (AST). When implementing a preprocessor, the user experiences the AST Query Mechanism as an interface to easily define and perform a subset of queries. This subset will typically be used to ask for a standard library list satisfying certain conditions, and it is possible to query for this in both a sub-tree of the AST and a list of SgNode pointers.

An example of a query is to ask for all Variable Declarations in the AST. The AST Query Mechanism introduces a method to easily ask the AST for Nodes satisfying conditions specified by the user, and also implements a set of predefined queries. The three different queries are the Name Query, Node Query and the Number Query. They are respectively returning a list of STL strings, a list of SgNode pointers and a list of integers. They all ask questions on all nodes in a sub-tree of the AST but are interested in different kinds of information. Because the AST Query Mechanism can query both a sub-tree of the AST and a list of SgNodes, it is often useful to combine the different queries to form an algorithm. For instance, one may ask the Node Query to create a list of all the Class Declarations in the AST, and another may ask the Name Query to query this list and return the names of all the classes.

The C/C++ Graphing Mechanism is an easy way to statically graph C/C++ data structures and provide a way to automatically generate code to dynamically graph the initialized data structures of a computer program at runtime. A graph of a computer program is a picture representing the relationship between objects and elements in the objects. For the user, the C/C++ graphing mechanism is a preprocessor that outputs a graph of the static data structure and the code needed at runtime to dynamically graph the code. When statically graphing the data structure, the data elements, such as variable declarations and class declaration, are graphed, and the static inheritance relationship is graphed. Dynamic graphing also graphs the data elements but is not graphing the inheritance relationship. The relationships that are graphed at runtime are the objects initialized to a pointer or reference and the values assigned to the defined variables. A dynamic graph is a snapshot of the memory allocated by the program at a specific time in the execution.

The AST Query Mechanism and the C/C++ Graphing Mechanism has successfully added value to the ROSE preprocessor project. Through these tools, the library developer can increase productivity while writing a preprocessor and debugging a program.

TOC

Implementation of OpenMP Support in the gFortran Compiler

Student **Elmer Salazar,** California State University, Stanislaus

Mentor **Mike Kumbera,** DCOM

OpenMP is a specification for a collection of compile-time directives and runtime functions in C and FORTRAN. This collection serves as a parallel programming interface. By using a standardized specification, the platform-specific details, as well as certain complexities of parallelism, can been hidden from the programmer. This allows for cleaner, more portable code, and compiling as parallel code is optional. If the programmer chooses to not compile as parallel code, the directives are treated as comments (ignored). This project's goal is to implement support for the OpenMP standard into the gFortran compiler. The gFortran compiler is a FORTRAN compiler that is part of the GNU GCC package. By implementing this standard into gFortran, programmers have access to OpenMP features, as well as the ability to see the intermediate code, including the transformations made by the compiler to handle parallelism.

Extending ReiserFS for Automatic File Queuing

Student

Jennifer Sirp, California State University, Sacramento

Mentor

Terry Brugger, NAIC

Typically, file systems are not concerned with directory entry organization. Tools like *ls* and *sort* are used to organize directory contents, but these tools perform linear operations that are costly when directories grow large. If files were maintained in the order they were written and stored in a priority queue, then a simple system call could return the "next file" to be processed. This summer, the benefit of adding automatic file queuing functionality to ReiserFS was explored.

A priority queue was implemented as an array of linked lists and maintained in kernel memory. As files are written on the partition, their names are examined to determine their priority and are then stored in the queue for rapid access. Inserts and retrievals are performed at a constant rate and priority queues exist for each directory that is created on the partition. Statistics on the queues are published through the proc File System, and access to the next file is achieved through a call to the *ioctl* function.

Tests were conducted on the new file system. A significant performance increase was achieved over prior user space applications that relied on multiple system calls and directory listings. Currently, work is being pursued on preserving the queue contents in hidden files so the data can be recovered in the event of system failure.

Efficient Schemes for Hyperbolic Systems with Stiff Relaxation Source Terms

Student

Yoshifumi Suzuki, University of Michigan, Department of Aerospace Engineering

Mentor

Jeffrey Hittinger, CASC

The objective of this study was to understand and develop numerical schemes that accurately and inexpensively solve hyperbolic systems with stiff relaxation source terms, systems that commonly arise in non-equilibrium hydrodynamic and radiation hydrodynamic problems. The generalized hyperbolic heat equations (GHHE) were used as a model system to investigate

$$\partial_t u + \partial_x v = 0$$
$$\partial_t v + \partial_x u = -\frac{1}{\varepsilon}(v - ru)$$

numerical schemes. The desired property of numerical methods for this problem is asymptotic preservation (AP), that is, the ability to accurately compute the diffusion limit when the relaxation process is underresolved. For this property, the scheme must account for the coupling between the advection and relaxation operators.

I briefly investigated the gas-kinetic scheme developed by Kun Xu. The idea of this BGK scheme is to evolve an approximation of the particle distribution function in each cell from which the macroscopic flow variables are updated. This scheme can accurately compute the Navier-Stokes limit by solving a Euler-like system provided that the distribution can support departures from equilibrium, but it is not clear how to extend this approach to more general systems.

Next, I investigated a particular semi-implicit Discontinuous Galerkin (DG) method [1] that has been shown to have the AP property in 1D. Accuracy in the diffusion limit is maintained because the solution slope is evolved directly with equations that account for the effect of the source term. I verified numerically that this scheme has the AP property in 1D by conducting convergence studies that compared to an exact solution. However, the AP property may be lost in higher dimensions. I have written a 2D DG code for a 2D extension of the GHHE system to investigate this issue, but I have not yet completed the convergence study. Investigating the 2D AP property of the DG scheme both numerically and analytically is the basis for future work.

Publications

[1] Lowrie, R.B. and J.E. Morel, "Methods for hyperbolic systems with stiff relaxation," *Int. J. Numer. Meth. Fluids* **40**, 2002, pp. 413-423.

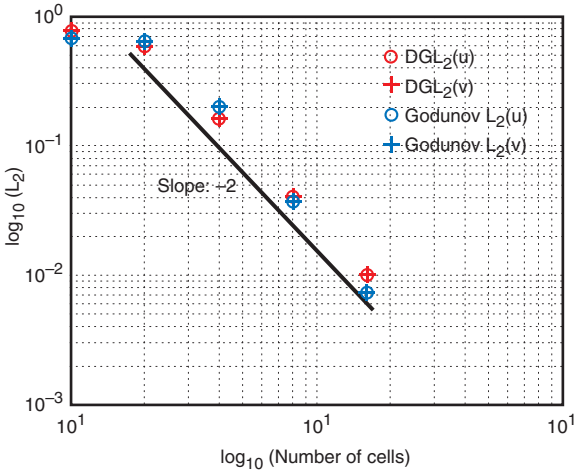


Figure 1. Grid convergence study of the DG method and the finite-volume Godunov method in the advection limit $(\varepsilon = 10^3, r = 0.5)$ solving the GHHE. This figure shows that both methods have second-order accuracy in this limit.

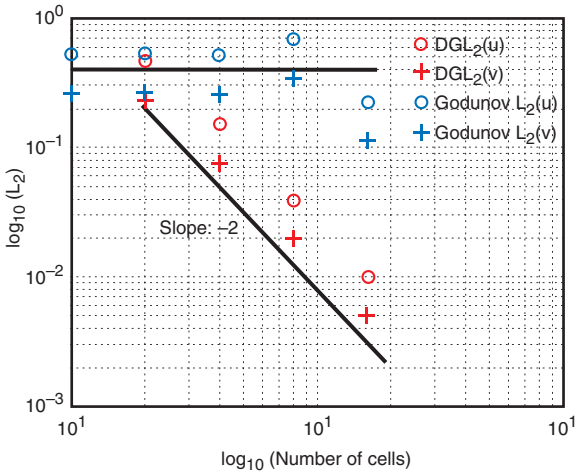


Figure 2. Grid convergence study of the DG method and the finite-volume Godunov method in the diffusion limit $(\varepsilon = 10^{-5}, r = 0.5)$ solving the GHHE. The DG method maintains second-order accuracy in this limit (AP property) unlike the Godunov method.

Streaming Computation of Structural Graphs

Student

Valerie Szudziejka, University of California, Davis

Mentor

Valerio Pascucci, CASC

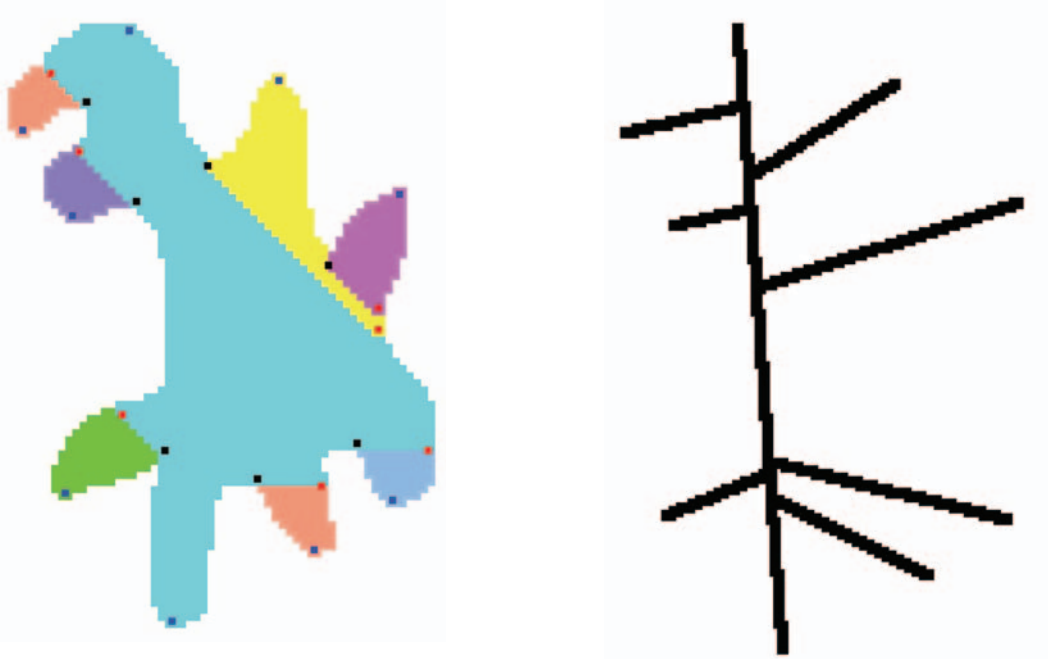
Contour trees are useful for describing data simply while preserving extrema. We have implemented an efficient algorithm for constructing contour trees on two-dimensional data sets. This approach is easy to implement, and unlike previous algorithms, does not require us to sort any data, and scales to large data sets.

Our algorithm is expected to run in linear time in the number of critical points. One difference between our algorithm and previous algorithms is that we follow ascending edges of a mesh rather than following the isolines. It is simpler to label regions but more complicated to connect up the contour tree.

The colored figure below is a simple region where height corresponds to its y-coordinate. Maxima and minima are blue and saddles are black. Red points are not true extrema, they are 'starting points' that arise because our algorithm follows ascending edges of the mesh only.

Ordinarily, branches of a contour tree connect at a saddle point. In our method, branches connect along the line that connects the false extrema to its saddle. The final contour tree is equivalent—the stick picture on the right.

Future work includes extending this algorithm to three-dimensional data sets and adapting an out-of-core hierarchical representation of the data set such as the Z-order space-filling curve.



The colored image (left) is encoded with Z-order space-filling curves. The corresponding structural graph (right) is called a Contour Tree and is obtained by contracting each region of a distinct color to an arc in the tree. The computation is redesigned in a non-conventional way to achieve high performance in a streaming infrastructure.

TOC

Parallel AMG for Systems of PDEs

Student **Ryan Szypowski,** University of California, San Diego

Mentor **Ulrike Yang,** CASC

Algebraic multigrid (AMG) is a linear solver applicable to a wide variety of problems. However, when the linear system to be solved is derived from a system of coupled partial differential equations (PDEs), standard AMG often performs unsatisfactorily.

In such cases, it is necessary to augment the standard AMG scheme using knowledge of which unknowns are associated with a single node in the discretized PDE. One approach, known as the unknown approach, is to ignore connections between the different functions and coarsen them separately. Another approach, known as the nodal approach, is to categorize nodes of the discretization—as opposed to unknowns—as coarse or fine nodes, thereby keeping the coupled structure through all levels of the multigrid hierarchy.

The nodal approach has been studied in past work, and for small elasticity problems, it was seen to show an improvement over standard AMG. To test this method on larger problems, it was necessary to extend code previously written to work in the parallel version of Hypre. A large amount of code was added to Hypre to accommodate parallel block-structured sparse matrices. Also, special interpolation and relaxation routines were developed for use with the block-sparse matrices.

Although the code is not yet complete, we have hopes that this method will work well for a wide range of elasticity problems.

Computation of Cellular Detonation

Student

Brian Taylor, University of Illinois, Urbana-Champaign

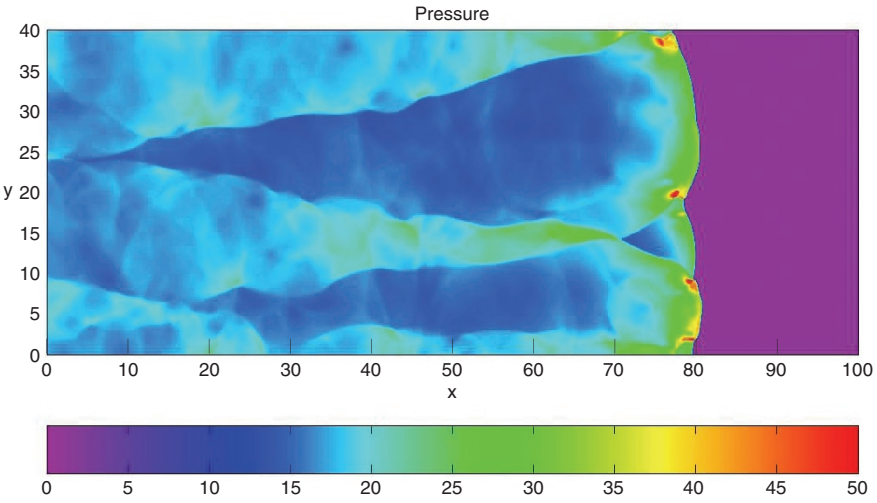
Mentor

Bill Henshaw, CASC

A detonation is a region of rapid chemical reaction in which a very strong shock wave is sustained by the reaction heat release. Detonations can occur in explosives that are premixed gases, liquids, and solids. The nominal speed of a steady detonation wave is experimentally found to be close to the Chapman-Jouguet (CJ) detonation wave speed, which corresponds to a case where the state at the end of the reaction zone is sonic relative to the steady frame. Typically, the lead shock dynamics are influenced only by a finite region, mostly in the reaction zone.

Detonations are subject to strong multidimensional, cellular, and chaotic instabilities associated with shock-shock dynamics. Shock collisions may generate disturbances in the reaction zone. Typically, two portions of a detonation shock that collide will generate an oblique shock interaction. The oblique shock can generate reflected shocks and slip lines, which induce short-lived Kelvin-Helmholtz instabilities in the region near the detonation front. In two dimensions, the steady detonation is unstable and develops a cellular structure.

My work this summer involved simulating cellular detonation with a simple one-step reaction mechanism. I worked with a code developed at UIUC and with OverBlown, an Euler solver written using the Overture PDE framework. Primarily, I was interested in applying adaptive mesh refinement (AMR) to this problem using OverBlown, as this is a feature that we would like to incorporate into the code my research group at school is developing. I also spent quite a bit of time looking at how object-oriented programming can best be used in numerical programs. OverBlown/Overture is a large object-oriented C++ code. The design philosophy underlying it was quite instructive to learn about, especially the ways in which C++ is used to organize and encapsulate data with Fortran implementing the core numerics.



Cellular Detonation Pressure Profile

Vulnerability Tracking Database 2.0

Student

Rober Taylor, Northern Arizona University

Mentor

Jody Malik, CSP

The Vulnerability Tracking Database (VTDB) 2.0 project involved a rapid-development parallel redesign and upgrade of the current vulnerability tracking database and database interface scripts. In order to fully develop the new version of the database and interface within the short time period, I worked directly with Barry Dahling (CSP). Mr. Dahling was responsible for the major database modifications and redesign, while I was responsible for the interface and database interactions.

The current interface, implemented in single 3000-line Perl file, was redesigned and rewritten to create an easy-to-maintain PHP file hierarchy. For the interface, a new layout using Cascading Style Sheets (CSS) was created to allow for a simple, yet elegant, HTML interface to the new database. In addition to the redesign, many requested components were implemented in the interface. A knowledge base was added to the interface to provide vulnerability solutions and comments. The single load was upgraded to a bulk loader to increase productivity by reducing the number of repetitive tasks. Last, a new host report was created to provide vulnerability and comment history of a given host.

One of the main concerns with the current interface is access control. Therefore the VTDB 2.0 and accompanying interface were designed with access control in mind. Finer granularity was added to the VTDB, allowing for the designation of read and read-write users, along with finer directorate control. For the interface, I implemented an authorization component to control access to the interface and perform directorate control and checks.

Global Analysis of the ROSE Infrastructure

Student

Nils Thuerey, University of Erlangen-Nuerenberg, Germany

Mentor

Dan Quinlan, CASC and **Jody Malik,** CSP

The ROSE framework allows the transformation of C++ source code taking into account the semantics of higher-level abstractions. As its use is intended to be similar to a standard compiler, the need for global analysis results for a project arises. Hence, the first objective was to introduce a database that is able to store that information to ROSE. Furthermore, callgraph and classhierarchy traverals were chosen to test the database and simplify further program analysis within ROSE. As the callgraphs of typical programs turned out to be large, a tool for interactive visualization of graphs with many nodes and edges was written.

For the database management system, MySQL was chosen due to its powerful features and its availability as Open Source software. Additionally, the C++ interface to the database MySQL++ is used. The callgraph and classhierarchy generation programs were implemented as ROSE preprocessors. Both graphs use the BOOST graph library to simplify common tasks. While the classhierarchy generation was straightforward, callgraphs in C++ have to handle several language specific issues.

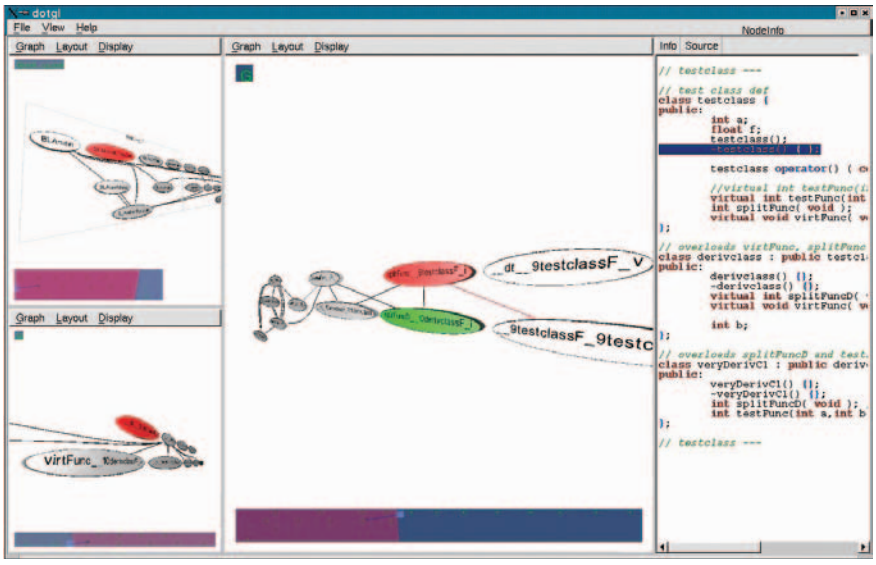
Function pointers and virtual functions make callgraph generation for C++ difficult. Virtual functions were handled by using the classhierarchy to identify virtual functions and possible calls to inheriting classes redefining these functions. For exchanging graphs with other programs, the .dot format from the GraphViz package is used within ROSE. However, as the Graphviz tools were not able to conveniently display larger graphs, a tool using the hardware acceleration of modern graphics cards to display and traverse graphs was written (dotgl). It uses the Graphviz development libraries to layout graphs and generate positions and sizes of nodes, edges and labels in the graph. The graph is then displayed with OpenGL. Due to graphical optimizations, such as adaptive level of detail and view frustrum culling, even graphs with several thousand nodes can be displayed in realtime.

The database interface, callgraph and classhierarchy are part of the ROSE framework now and will be used in future program analysis. The dotgl graph visualization will be maintained as a sourceforge project. Future work on the callgraph will handle function pointers, and evaluate its use for the automatic generation of components.

Publications

D. Quinlan & M. Schordan, "ROSE: A Tool for building Source-to-Source Preprocessors," 2003.

Y. Chiricota & F. Jourdan & G. Melancon, "Software Components Capture Using Graph Clustering," 2003.



Three different callgraphs generated with ROSE and displayed with dotgl are shown. The sourcecode for the selected node of the center callgraph is shown in the panel to the right.

Code Validation Made Easy

Student **Peter Tipton,** University of Southern California

Mentor **Brandon Scott,** AX

Using DAKOTA (Design Analysis Kit for Optimization and Terascale Applications), users can perform a variety of studies to see how their code runs under different conditions. Among other things, users can run parameter studies, uncertainty quantification, and optimization. Using these tools, users can see how their code's output changes as their input parameters are varied, semi-automating the process of code validation.

The goal of this project is to take DAKOTA and be able to run it on an arbitrary code on an arbitrary machine. This should be done in such a manner that users will be able to analyze their code in the easiest, most powerful way possible. Our implementation of this technique is done with Python scripts.

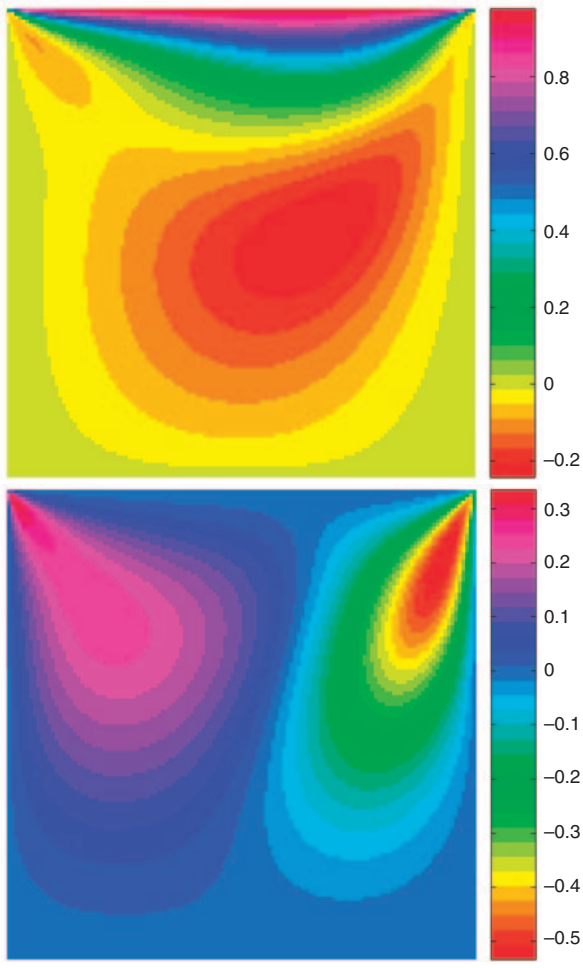
A MATLAB Implementation of Mixed Finite Element Method for Incompressible Newtonian Flows: Pseudostress-Velocity Formulation

Student **Chunbo Wang,** Purdue University

Mentor **Charles Tong,** CASC

We investigate the use of a mixed finite-element method based on the pseudostress-velocity formulation for simulating incompressible Newtonian fluid flows. Semi-implicit time discretization of the formulation leads to a decoupled system at each time step such that one can calculate the pseudostress first and then the velocity. Physical quantities such as stress, vorticity and pressure can all be computed algebraically from pseudostress. The use of Raviart-Thomas finite elements in discretizing the decoupled pseudostress equation gives a linear system of equations that can be solved efficiently by a multigrid-preconditioned conjugate gradient method.

We developed a MATLAB code to show this scheme applied to the Lid-Driven Cavity problem. We also verify the accuracy of this scheme numerically by considering a 2-D Stokes problem. A GMRES preconditioned with H(div) geometric multigrid method is also developed.



Plot of velocity of Lid-Driven Cavity problem

Newton–Krylov Methods for Expensive Nonlinear Function Evaluations

Student **Rebecca Wasyk,** Worcester Polytechnic Institute

Mentor **Carol Woodward,** CASC

Newton-Krylov methods have proven useful for solving large-scale nonlinear systems. An advantage of these iterative methods is that they do not require storage of the system Jacobian, but only require knowledge of how the Jacobian acts on a vector. A difference quotient evaluated at each linear iteration is often used to approximate this action without slowing the convergence rate of the method. For systems with expensive nonlinear function evaluations, however, the requirement of a function evaluation for each linear iteration can result in a very costly computation. The goal of this project is to explore convergence rates and time savings associated with using different approximations to the system function in the difference quotient.

Theoretical results developed in cooperation with Peter Brown, Homer Walker, and Carol Woodward indicate conditions on an approximation function that could be used instead of the full nonlinear function in the difference quotient without negatively impacting the local convergence rate. One such approximation that satisfies these conditions is approximating the most expensive parts of the nonlinear function with a linear approximation. Since making a linear approximation could be expensive, we also tested a method where the most expensive nonlinearities were lagged, so that they did not need to be recomputed in the difference quotient at each linear iteration.

These two approximations schemes were tested on several problems using the KINSOL solver in the SUNDIALS program, a suite of solvers developed at LLNL for solving nonlinear algebraic systems, ordinary differential equation systems, and differential-algebraic systems. For most of the problems tested, the linear approximation method did converge in the same number of iterations as were required when using the unmodified difference quotient, though in a few cases failures resulted when using this approximation. When the lagged approximation worked, we saw significant drops in the runtime in comparison to the unmodified method. As predicted by theory, the linear approximation had the same convergence properties as using no approximation and there were some noticeable time savings using this linear approximation on a few problems. However, none of the test problems had the type of expensive nonlinearities often seen in real application codes, where we would expect to see the most time savings.

Future work aims to develop some theoretical results on solving time-dependent problems using these approximations in the difference quotient. More tests are also planned on problems with more expensive nonlinearities, where the modifications made could really be beneficial in terms of runtime savings.

Metadata Management for Petabyte-Scale File Systems

Student **Sage Weil,** University of California, Santa Cruz

Mentor **Tyce McLarty,** CADSE

Previous simulation work validated many of the basic design points of a dynamic, distributed, subtree-based metadata server for large object-based file systems. Performance of the metadata server in file systems that decouple metadata from read and write operations is critical to overall system performance and scalability. Real-world performance of the design at scale remains speculative, due to the small scale of past simulation and the impracticality of simulating metadata operations at scale (an order of magnitude larger than existing Lustre installations).

To further study metadata performance issues, I have created a fully distributed prototype of the metadata server. The system is based on an asynchronous (but single-threaded) programming model based on C++ and STL, utilizing class-based continuations for delaying execution. Most of the basic functionality has been implemented, including a generic messaging layer (currently utilizing MPI), dynamic load balancing, synthetic client workload generation, and most major metadata operations needed to support the POSIX file system interface.

Information Leakage due to Geographic Properties of Internet Routing

Student **Daniel Wendlandt,** Stanford University

Mentor **Martin Casado,** ISCR

While it is commonly known that Internet routes operate with no restraint toward geopolitical boundaries, no empirical studies to date have explored how this notion of a network without borders impacts a nation-state's ability to limit access and control over its Internet traffic.

We sought to explore the issue of information leakage caused by geographic properties of Internet routing. We define information leakage to be when a route with both its source and destination in a single country uses infrastructure in another nation for transit. A better understanding to information leakage is crucial for national security concerns over controlling access to vital information and services running over the Internet. Our goal was to define a problem space, share data results, discuss underlying causes of information leaking paths, and explore diverse topics for future work in this area.

We devised a simple methodology for gathering introductory data on information leakage by using geolocation software to analyze traceroute paths between source/destination pairs within a single country. Our data results are compelling not only in that they are the first attempt to quantify the amorphous nature of information leakage but also because even our lower-bound results show that even for many highly advanced countries, the degree of information leakage is strikingly high. Additionally, our analysis shows that these circuitous paths are highly dependant on the type of service provider offering access to both the source and destination. We also explore how highly fiber-rich regions called "routing hubs" affect Internet routes.

TOC

Ameliorating the Performance Degradation of User-Defined Abstractions and Indirect Memory Accesses

Student **Brian White,** Cornell University

Mentor **Daniel Quinlan,** CASC

The complexity of physical simulations performed at LLNL demand high-level abstractions, such as vectors, tensors, and fields. While such abstractions facilitate program correctness, maintainability, and readability, they often inhibit performance because a compiler lacks sufficient semantic understanding to optimize them. For example, we encountered a field vector implementation based on a standard template library vector. Though the length of the field vector was known statically, it was not available to the compiler. Replacing the STL vector with scalars removed indirection inherent in the implementation and the overhead associated with its iterator and resulted in a 2x speedup.

Mesh-based codes exhibit indirect memory accesses, for example, as they visit nodes associated with an edge. Lab codes utilize iterators to mask such indirection. Unfortunately, the required indirection maps and iterator incrementing and dereferencing contribute to an excessive branch misprediction rate. One code suffered a 12.23% misprediction rate when measured on a Power3, achieving an anemic 0.88 instructions per cycle as a result on a machine with peak performance of 8 CPI (clocks per instruction). These unpredictable branches appear to be the performance bottleneck; a data-packing optimization that reduced L1 misses by 18% had no discernable effect on performance given the misprediction rate.

Despite the performance degradation induced by high-level abstractions, we believe they are necessary for effective programming. We seek to provide the performance benefit of more efficient, but less succinct, code through automatic transformation within the ROSE framework. Automating some of the above techniques will serve as the starting point for such future work.

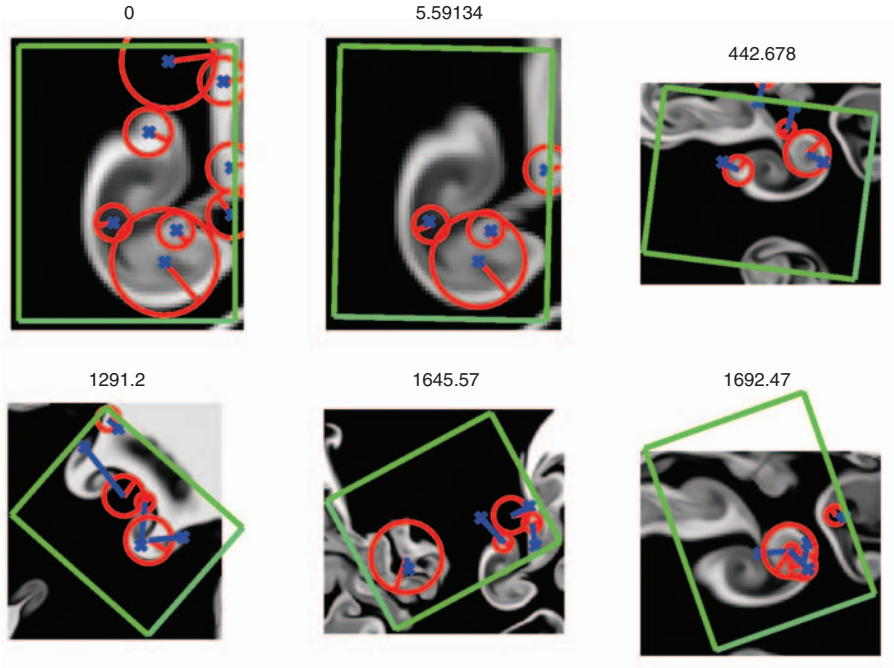
Matching Shapes Using Local Descriptors

Student **Ryan M. White,** Univeristy of Calilfornia, Berkeley

Mentor **Shawn Newsam,** CASC

We present a method for comparing shapes of grayscale images in noisy circumstances. By establishing correspondences in a new image with a shape model, we can estimate a transformation between the new region and the model. Using a cost function for deviations from the model, we can rank resulting shape matches. We compare two separate feature detectors—Scale Saliency and Difference of Gaussians. We show that this method is successful in comparing images of fluid mixing under anisotropic geometric distortions and additive Gaussian noise and that Scale Saliency outperforms the Difference of Gaussians.

This project is further described at the Sapphire Web site:
http://www.llnl.gov/casc/sapphire/sapphirehome.html



The green box in the top image represents the query region. The remaining images show the best matches in images that have been geometrically altered. The red circles are the features.

Additions to the ROSE Compiler Infrastructure

Student

Jeremiah Willcock, Indiana University

Mentor

Daniel Quinlan, CASC

My project this summer was to add to the ROSE project, a framework for creating source-to-source transformations for C++ programs. This project represents C++ programs as abstract syntax trees (ASTs) and provides features to parse C++ programs into ASTs, as well as to convert from ASTs back into C++ source code. I added several transformations on ASTs to the collection provided by ROSE.

One optimization I implemented was function inlining. This transformation replaces a call to a function with the body of the function, saving the overhead of the call, as well as enabling other optimizations on the combined source code. I also implemented two other optimizations using the ROSE framework—partial redundancy elimination and finite differencing. These standard compiler optimizations remove redundant computations and simplify complex computations. All three of these transformations are now included in the current version of ROSE, so ROSE users can also employ these transformations as part of their own optimizations.

Recently, I prototyped an interface between the Sage AST format used by ROSE and the Aterm file format for representing tree structures. This interface will enable the use of advanced rewriting tools, such as Stratego, within ROSE. In addition to my code additions to the ROSE project, I also contributed many bug reports and suggestions for improvement. Overall, I added several components to the ROSE project during my practicum this summer; these components will make ROSE more useful both for the Laboratory and for its external collaborators.

Shallow Water Equations on Curvilinear Grids

Student

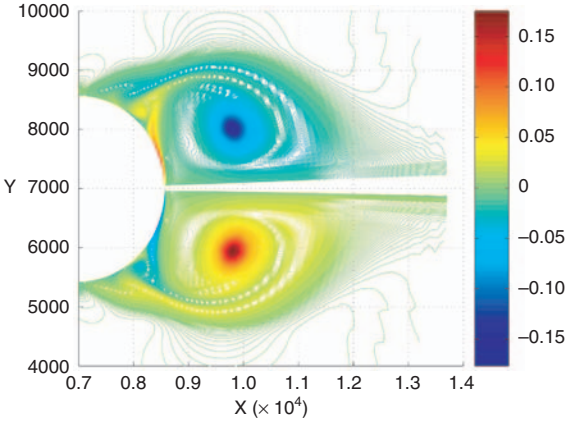
Suzanne Wingenter, San Diego State University

Mentor

Petri Fast, CASC

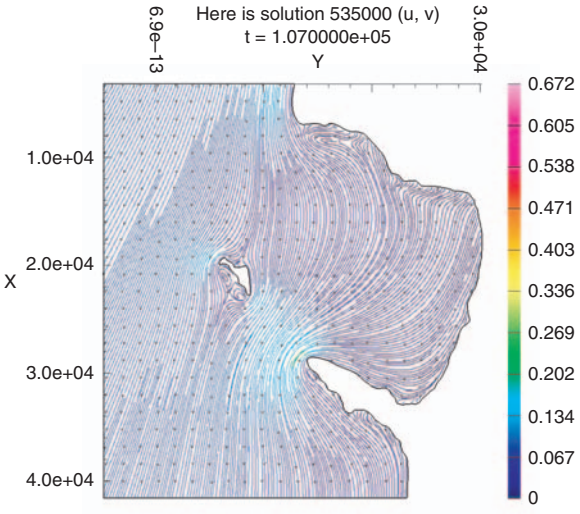
This project solves the curvilinear version of the shallow water equations as described by Borthwick and Barber. [1] The equations were solved using an Adams Predictor Corrector scheme that was coded in C, compiled and run locally on the student workstation. This version was used to validate the finite-difference scheme using Richardson convergence rate studies. Convergence rates indicate the scheme accurately solves the equations with second-order convergence.

Three more test cases were prepared to check logical flow behavior. The first involves the development of flow in a straight channel over a small seamount showing the effects of bottom bathymetry on the surface elevation. Second, flow through a generic river geometry tests the ability of the flow in a contorted geometry. And third, flow past an idealized island tests recirculation. All three test cases show excellent behavior.



Flow past an idealized island shows recirculation, vorticity plot.

The last part of the project reimplements the code in Overture using native operators. Several simplified examples of the Bahia de Todos Santos (Baja California, Mexico) were run, the first of which is a smooth, coarse grid of the bay that can be run quickly. The other two examples were a flow past the northern most entrance to the bay, and a flow past the island of the bay. The last simulation performed involved the full bay. The simulation is running currently and the behavior is still in its transient phase.



Transient behavior in Overture simulation of Bahia de Todos Santos, streamline plot showing beginning "steady state" tidal flow as the tide is starting to come in.

Publications

[1] A.G.L. Borthwick, RW. Barber, "River and Reservoir Flow Modeling using the Transformed Shallow Water Equations" *International Journal for Numerical Methods in Fluids*, 14:1193-1217, (1992).

A Hybrid Sort-First/Sort-Last Approach for Rendering Translucent Geometry in the VisIt Visualization Tool

Student

Christopher Wojtan, Georgia Institute of Technology (Fall 2004), University of Illinois, Urbana-Champaign (Fall 2000 – Spring 2004)

Mentor

Jeremy Meredith, DNT

The scientific visualization tool VisIt lacked any support for reading image formats as data sets. I implemented support for image file formats (PPM, BMP, JPEG, TIFF, etc.) within VisIt. The red, green, and blue channels within these images were treated as vector data and the pixel location was treated as the spatial location of the data. Then I created a way to plot the image in red, green, and blue colors, instead of using a variable color ramp to display each channel. VisIt did not handle translucency in two-dimensional data very efficiently, so I improved that, as well. The original scheme involved sorting all cells front to back, then calculating the transparency. Sorting was not necessary in two dimensions.

After this work, I began correcting the way VisIt displays translucent data when rendering in parallel. The original method assigned a group of data to each processor according to the data's location in world space. This led to problems with transparent data, since it needs to be sorted from front to back in image space. We decided to assign data to processors according to locations in image space in order to correctly perform this scalable rendering. Since the data was originally assigned incorrectly for our purposes, the processors needed to redistribute the data. I wrote code for each processor to transform its data into image space and redistribute everything correctly in parallel. Now the data is correctly rendered in each region of the screen, and we only need to consider clipping everything to finalize this scheme.

Cache Coherent Mesh Layout

Student **Sung-Eui Yoon,** University of North Carolina at Chapel Hill

Mentor **Peter Lindstrom,** CASC

Recently, streaming meshes have been introduced to efficiently perform out-of-core processing of gigantic meshes. The mesh is constructed such that memory requirements are minimized during processing the mesh following a streaming (or stored) order. Moreover, seamless connectivity of currently processing elements is provided. This streaming mesh greatly improves out-of-core performance of applications that do not require their own specific processing order. However, many applications require their own specific processing order, which is different from a stored streaming order.

To improve out-of-core performance of those applications, we need to have a mesh layout, which is stored in a specific order and, more importantly, maintain locality of mesh elements for cache coherent access pattern.

There is plenty of literature available on layouts of structured meshes (e.g., uniform grid). Classic space-filling curves have been used to compute ordering of mesh elements to maintain locality. However, embedding a space-filling curve into a uniform geometric structure can deteriorate quality of locality in unstructured meshes. To ameliorate this problem, we figured out an algorithm computing the ordering based on construction of classic space curves on unstructured meshes. Moreover, to quantify locality of the ordering, we analytically compute a probability of a cache miss according to edge length given cache parameters.

In the future, we plan to design a cache-oblivious algorithm that does not require exact cache parameters. This work will be carried out at University of North Carolina with Peter Lindstrom.

TOC

A Visual Interface for the Promoter Identification Workflow

Student **Beth Yost,** Virginia Polytechnic Institute and State University

Mentor **Terence Critchlow,** CASC

The next step in analyzing the human genome involves understanding gene regulation. Genes are turned on or off based on a complex series of mechanisms. These controls involve the promoter region of a gene, typically located right before the start of transcription. In this region, transcription factor binding sites can be found. Biologists determine promoter modules based on the frequency of these sites and the distance between them. By comparing these factors across multiple sequences, it is possible to determine other genes that might have a similar function.

Identifying these promoter regions is a difficult task. The Promoter Identification Workflow (PIW) automates this process, but a visual interface designed specifically for the biologist was lacking. We have begun developing an interactive visual interface for the PIW. This interface provides biologists with a quick overview of the results. Included in the overview is the ClustalW alignment of the sequences, shaded regions of those sequences representing the locations where there is homology across sequences, and colored boxes showing the locations of transcription factor binding sites (determined by TransFAC). A consensus sequence is calculated and is always shown at the top of the overview. There is a sliding window that shows details of 50 base pairs at a time. There are also filters based on the names and frequencies of transcription factor binding sites. Using the PIW with the added visual interface can not only speed the work of the biologist, but also provide them with additional insight.